

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 18:52:11 ; Search time 29 Seconds
(Without alignments)
2337.569 Million cell updates/sec

Title: US-09-870-406a-33

Sequence: 1 MALSMERFSGSCFKAPNP.....SLGKSYETRPMDLTGQFIG 329

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPREMBL_21.*

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rviro:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB ID	Description
1	329	100.0	329	10	09SR43
2	143	43.5	273	10	09SR72
3	9	2.7	240	10	08R2B6
4	8	2.4	371	17	0970H0
5	8	2.4	491	10	09FZC2
6	8	2.4	751	16	08RDB0
7	8	2.4	786	10	09CTE7
8	7	2.1	59	1	09YCA3
9	7	2.1	59	11	09C0Y3
10	7	2.1	88	2	09R0H4
11	7	2.1	94	15	077822
12	7	2.1	107	1	P94122
13	7	2.1	109	6	09N0W5
14	7	2.1	124	2	09ARF5
15	7	2.1	124	2	08VSP8
16	7	2.1	124	2	08VSE7

17	7	2.1	126	16	P96817	P96817 mycobacteri
18	7	2.1	136	16	09FBW2	09FBW2 streptomyc
19	7	2.1	149	6	095UP1	095UP1 macaca fasc
20	7	2.1	158	12	091SF2	091SF2 trichoplus1
21	7	2.1	158	17	058392	058392 pyrococcus
22	7	2.1	164	2	053026	053026 enterococcu
23	7	2.1	167	2	08RLX5	08RLX5 lactobacill
24	7	2.1	193	6	097940	097940 sus scrofa
25	7	2.1	213	1	093657	093657 methanosarc
26	7	2.1	213	17	08TTB0	08TTB0 methanosarc
27	7	2.1	213	17	08TN69	08TN69 methanosarc
28	7	2.1	218	17	08T871	08T871 methanosarc
29	7	2.1	224	16	08XNK2	08XNK2 clostridium
30	7	2.1	229	12	041089	041089 paramecium
31	7	2.1	244	11	09CUZ0	09CUZ0 mus musculu
32	7	2.1	248	9	08SBY3	08SBY3 streptococc
33	7	2.1	249	2	09FD15	09FD15 streptomyc
34	7	2.1	253	16	091018	091018 streptomyc
35	7	2.1	256	16	08ZEH4	08ZEH4 yersinia pe
36	7	2.1	256	16	08XXG7	08XXG7 ralsconia s
37	7	2.1	257	16	092N06	092N06 rhizobium m
38	7	2.1	262	16	069841	069841 streptomyc
39	7	2.1	277	8	08WHY1	08WHY1 psittolum nu
40	7	2.1	283	16	067855	067855 aquifex aeo
41	7	2.1	287	16	092605	092605 listeria in
42	7	2.1	287	16	08Y312	08Y312 listeria mo
43	7	2.1	292	13	09PSA4	09PSA4 gallus gall
44	7	2.1	294	16	09KN35	09KN35 vibrio chol
45	7	2.1	304	16	09R204	09R204 delnoccoccus

ALIGNMENTS

RESULT 1
ID 09SR43 PRELIMINARY; PRT: 329 AA.
AC 09SR43:2060 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE F3124.1 protein (PHYTOCHROMOBILIN synthase HY2 protein).
GN F3124.1 OR HY2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Kosteidae;
OC Eustoides II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Ronning C.M., Koo H., Fujii C.Y., Uterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Niernan W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC F3124 genomic sequence.";
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=21124703; PubMed=11226195;
RA KOEHL T., Mukougawa K., Frankenberg N., Masuda M., Yokota A.,
RA Lagarias J.C.;
RT "The Arabidopsis HY2 Gene Encodes Phytochromobilin Synthase, a
RT Ferredoxin-Dependent Biliverdin Reductase.";
RL Plant Cell 13:425-436(2001).
DR EMBL; AC011436; AAF14017.1; -;
DR EMBL; AB045112; BAB3374.1; -;
SQ SEQUENCE 329 AA; 38129 MW; 98727CD1778D973C CRC64;
Query Match 100.0%; Score 329; DB 10; Length 329;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1, MALSMERFSGSCFKAPNPVLISAPNKNIFLRRKRKRLRVAVSKEFAESALE 60

```

Db 1 MALMEGFSIGSCFAPNPVILASPKINFTLRKKKFLRVSAYSKFASALE 60
QY 61 ETRRIYLEPSHLOEYSSMTGLDCKTELQMLAKRSKIRLSMAIENETMOVDFAG 120
Db 61 ETRRIYLEPSHLOEYSSMTGLDCKTELQMLAKRSKIRLSMAIENETMOVDFAG 120
QY 121 MEPEYTPFCANFTSTNVNIVLDPHQLTDQTDYQDKYKINSIYHKAETFPW 180
Db 121 MEPEYTPFCANFTSTNVNIVLDPHQLTDQTDYQDKYKINSIYHKAETFPW 180
QY 181 GSKLTGESIKFSPVMMTRSSSKERKALFSAFLEYQALFMTIOVREEMPSHVA 240
Db 181 GSKLTGESIKFSPVMMTRSSSKERKALFSAFLEYQALFMTIOVREEMPSHVA 240
QY 241 NCEAQHKTLYMRAQKDPGHLKRLVGEAKAEKELLROFLNGVDLSTKTFIDYPEYOT 300
Db 241 NCEAQHKTLYMRAQKDPGHLKRLVGEAKAEKELLROFLNGVDLSTKTFIDYPEYOT 300
QY 301 EDGTVDKRSIIIGKSYETSPMDLNGOFTG 329
Db 301 EDGTVDKRSIIIGKSYETSPMDLNGOFTG 329

```

RESULT 2

```

ID 09SS72 PRELIMINARY; PRT; 273 AA.
AC 09SS72;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DE 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
GN MZB10.18 protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eustoids II; Brassicales; eudicotyledons; core eudicots; Rosidae;
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Kohnig C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Niernan W.C., Fraser C.M.,
RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC009326; AAD56331.1;
SQ SEQUENCE 273 AA; 31909 MW; 06E94272646E87 CRC64;

```

Query Match
Best Local Similarity 100.0%; Score 143; DB 10; Length 273;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 113 QVDFAGMEPEYTPFCANFTSTNVNIVLDPHQLTDQTDYQDKYKINSIYH 172
Db 74 QVDFAGMEPEYTPFCANFTSTNVNIVLDPHQLTDQTDYQDKYKINSIYH 172
QY 173 KYAEFFPMGKLTGESIKFSPVMMTRSSSKERKALFSAFLEYQALFMTIOVRE 133
Db 134 KYAEFFPMGKLTGESIKFSPVMMTRSSSKERKALFSAFLEYQALFMTIOVRE 133
QY 233 MEPSHVANQEAQHKLTLMRAQ 255
Db 194 MEPSHVANQEAQHKLTLMRAQ 216

```

RESULT 3

```

ID 08R2B6 PRELIMINARY; PRT; 240 AA.
AC 08R2B6;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DE 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DR 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

```

```

DE B147A04.9 protein.
GN B147A04.9..
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT clone:B147A04.;
RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003735; BAB6192.1;
SQ SEQUENCE 240 AA; 27459 MW; 7C3433A04E5FE9E2 CRC64;

```

Query Match

Best Local Similarity 100.0%; Score 9; DB 10; Length 240;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 142 IVVLDLNP 150
Db 109 IVVLDLNP 117

```

RESULT 4

```

ID 097UHO PRELIMINARY; PRT; 371 AA.
AC 097UHO;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DE 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
GN SSO3049.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RA MEDLINE-21332296; PubMed-11427726;
RA She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Aweyer M.J., Chan-Welher C.C.-Y., Fletcher C., Gordon P.M.K.,
RA De Moers A., Erasmo G., Fiercher C., Clausen I.G., Curtis B.A.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Garret R.A., Ragan M.A., Sensen C.W., Duguet M., Gasterland T.,
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus p2."
RT Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AE006896; AAK43149.1;
DR InterPro; IPR000683; GRO_IDH_MOCA.
DR Pfam; PF01408; GRO_IDH_MOCA.1.
DR Pfam; PF02894; GRO_IDH_MOCA.C.1.
SQ SEQUENCE 371 AA; 41760 MW; DAAE16466D530E8 CRC64;

```

Query Match

Best Local Similarity 100.0%; Score 8; DB 17; Length 371;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 272 KELLRDFL 279
Db 318 KELLRDFL 325

```

RESULT 5

```

ID 09FZC2 PRELIMINARY; PRT; 491 AA.
AC 09FZC2;

```

DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE TIK7 26 protein.
 GN TIK7.26.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Sakano H., Liu S.X., Yu G., Lee J.M., Lenz C., Pham P., Toriumi M.,
 RA Chin C., Chlou J., Chol E., Chung M., Gonzalez A., Hong B., Liu A.,
 RA Vaysberg M., Altafi H., Brooks S., Buehler E., Chao Q., Conn L.,
 RA Conway A.B., Hansen N.F., Johnson-Hopson C., Khan S., Kim C., Lam B.,
 RA Miranda M.K., Nguyen M., Palm C.J., Shin P., Southwick A., Davis R.W.,
 RA Ecker J.R., Federspiel N.A., Theologis A.;
 RT "The sequence of BAC TIK7 from Arabidopsis thaliana chromosome 1.";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC013427; AAF98579.1;
 SO SEQUENCE 491 AA; 52832 MW; 147A6584A8496A27 CRC64;

Query Match 2.4%; Score 8; DB 10; Length 491;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 274 LLRDFLN 281
 |||||
 DB 290 LLRDFLN 297

RESULT 6

O8RDB0 PRELIMINARY; PRT; 751 AA.
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hydrogenase maturation factor.
 GN HEP OR TTE0131.
 OS Thermanaerobacter tengcongensis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
 OC Thermanaerobacteriales; Thermanaerobacteriaceae; Thermanaerobacter.
 OX NCBI_TaxID=119072;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MB4T / JCM11007;
 RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
 RA Chen Y., Xue Y., Xu Y., Lei X., Huang L., Dong X., Ma Y., Ling L.,
 RA Tan H., Chen R., Wang J., Yu J., Yang H.;
 RT "A complete sequence of T. tengcongensis genome.";
 RL Genome Res. 12:689-700(2002).
 DR EMBL: AE012986; AAM3435.1;
 KW Complete proteome.
 SO SEQUENCE 751 AA; 85546 MW; 1CEDA90F16E6AF7 CRC64;

Query Match 2.4%; Score 8; DB 16; Length 751;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 TLRRKKR 41
 |||||
 DB 240 TLRRKKR 247

RESULT 7

O9C7E7 PRELIMINARY; PRT; 786 AA.
 AC O9C7E7;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE Hypothetical 88.6 kDa protein.
 GN F1K9.16.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewart K.,
 RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujil C.Y.,
 RA Gali J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marzilli A.,
 RA Millscher J., Miranda M., Nguyen M., Nieman W.C., Osborne B.I.,
 RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shin P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 thaliana.";
 RT Nature 408:816-820(2000).
 DR EMBL: AC069471; AAG51492.1;
 SO SEQUENCE 786 AA; 88637 MW; 0F89AB3BC352CE7 CRC64;

Query Match 2.4%; Score 8; DB 10; Length 786;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 267 GEAKAKEL 274
 |||||
 DB 448 GEAKAKEL 455

RESULT 8

O9YG3 PRELIMINARY; PRT; 59 AA.
 AC O9YG3;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Dimethylamine corrinoid protein Mbc (Fragment).
 GN Mfbc.
 OS Methanosarcina thermophila.
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TM-1;
 RA MEDLINE=97156682; PubMed=9006042;
 RA Ferguson D.J. Jr., Krzycki J.A.;
 RT "Reconstitution of trimethylamine-dependent coenzyme M methylation
 with the trimethylamine corrinoid protein and the isozymes of
 methyltransferase II from Methanosarcina barkeri.";
 RL J. Bacteriol. 179:846-852(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TM-1;
 RA MEDLINE=20225851; PubMed=10762254;
 RA Paul L., Ferguson D.J. Jr., Krzycki J.A.;
 RT "The trimethylamine methyltransferase gene and multiple dimethylamine

RT methyltransferase genes of Methanosarcina barkeri contain in-frame and
 RT read-through amber codons."
 RL J. Bacteriol. 182:2520-2529(2000).
 DR EMBL: AF153452; AAB38788.1;
 DR InterPro: IPR003312; B12-binding.
 DR Pfam: PF02310; B12-binding; 1.
 FT NON_TER 1
 SO SEQUENCE 59 AA; 6257 MW; 479E0F9115B5ECF CRC64;

Query Match 2.1%; Score 7; DB 1; Length 59;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 AKAKELL 275
 DB 52 AKAKELL 58

RESULT 9

09CUC3 PRELIMINARY; PRT; 59 AA.
 AC 09CUC3;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DE 3230402N03Rik protein (Fragment).
 GN MAPKrip2 OR 3230402N03Rik.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RP [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=HIPPOCAMPUS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsi G.,
 RA Blake J., Boitelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Humé D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whitaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hvasizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL: AK013573; BAB28912.1;
 DR MGD: MGI:1926555; Mapk8ip2.
 DR InterPro: IPR000050; PTD_domain.
 DR Pfam: PF00640; PTD; 1.
 FT NON_TER 1
 SO SEQUENCE 59 AA; 6877 MW; 35C6B6E84D2FC2C CRC64;

Query Match 2.1%; Score 7; DB 1; Length 59;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 214 AFLERYQ 220
 DB 38 AFLERYQ 44

RESULT 10
 09R0H4

ID 09R0H4 PRELIMINARY; PRT; 88 AA.
 AC 09R0H4;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 GN Stage III sporulation protein J homolog (Fragment).
 OS Listeria monocytogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Listeriaceae; Listeria.
 OX NCBI_TaxID=1639;
 RP [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=EGD;
 RX MEDLINE=20208566; PubMed=10746777;
 RA Milohanic E., Fron B., Berche P., Gaillard J.L.;
 RT "Identification of new loci involved in adhesion of Listeria
 RT monocytogenes to eukaryotic cells."
 RL Microbiology 146:731-739(2000).
 DR EMBL: AF104228; AAF04771.1;
 DR InterPro: IPR001708; 60kDa_innermed.
 DR Pfam: PF02096; 60kDa_imp; 1.
 FT NON_TER 1
 FT NON_TER 1
 SO SEQUENCE 88 AA; 10281 MW; F021AF158CEBD2BC CRC64;

Query Match 2.1%; Score 7; DB 2; Length 88;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 LOEKYSS 79
 DB 19 LOEKYSS 25

RESULT 11
 077822 PRELIMINARY; PRT; 94 AA.
 AC 077822;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 GN Envelope protein (Fragment).
 OS Human immunodeficiency virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RP [1]
 RC SEQUENCE FROM N.A.
 RX MEDLINE=95191002; PubMed=7884875;
 RA Mulder-Kampinga G.A., Simonon A., Kulken C.L., Dekker J.,
 RA Scherpbier H.J., de Perre P., Boer K., Goudsmid J.;
 RT "Similarity in env and gag genes between genomic RNAs of human
 RT immunodeficiency virus type 1 (HIV-1) from mother and infant is
 RT unrelated to time of HIV-1 RNA positivity in the child."
 RL J. Virol. 69:2285-2296(1995).
 DR EMBL: Z47830; CAA87845.1;
 DR InterPro: IPR000777; GP120.
 DR Pfam: PF00516; GP120; 1.
 DR AIDS; Coat protein; Glycoprotein.
 FT NON_TER 1
 FT NON_TER 1
 SO SEQUENCE 94 AA; 10793 MW; E80573E5F9E2CC3D CRC64;

Query Match 2.1%; Score 7; DB 15; Length 94;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 HLOEKYS 78
 DB 80 HLOEKYS 86

RESULT 12

P94122 PRELIMINARY; PRT; 107 AA.
 AC P94122; (TREMBlrel. 03, Created)
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Hypothetical 12.0 kDa protein.
 GN ORF3
 OS Acidianus ambivalens (Desulfurolobus ambivalens).
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 CC Acidianus.
 OX NCBI_TaxID=2283;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DSM 3772;
 RX MEDLINE-971175566; PubMed-9023221;
 RA Pursche W.G., Schmidt C.L., Petersen A., Schaefer G.;
 RT "The Terminal Quinol Oxidase of the Hyperthermophilic Archaeon
 RT Acidianus ambivalens Exhibits a Novel Subunit Structure and Gene
 RT Organization";
 RL J. Bacteriol. 179:1344-1353(1997).
 DR EMBL; Y09614; CAA70829.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 107 AA; 11955 MW; BFC3623DB2B62878 CRC64;

Query Match 2.1%; Score 7; DB 1; Length 107;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 205 KEKHAL 211
 |||||
 DB 87 KEKHAL 93

RESULT 13

O9NOM5 PRELIMINARY; PRT; 109 AA.
 AC O9NOM5; (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Anti-human A33 light chain variable region (Fragment).
 OS Oryctolagus cuniculus (Rabbit).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-20250927; PubMed-10788485;
 RA Rader C., Ritter G., Nathan S., Elia M., Gout I., Jungbluth A.A.,
 RA Cohen L.S., Welt S., Old L.J., Barbas C.F. III;
 RT "The rabbit antibody repertoire as a novel source for the generation
 RT of therapeutic human antibodies";
 RL J. Biol. Chem. 275:13668-13676(2000).
 DR EMBL; AF245502; AAF68449.1; -;
 DR HSSP; P80362; 1MTL;
 DR InterPro; IPR003006; I9_MHC.
 DR InterPro; IPR003596; I9_V.
 DR Pfam; PF00047; I9; 1.
 DR SMART; SM00406; IGV; 1.
 FT NON_TER 1
 FT TER 1
 SQ SEQUENCE 109 AA; 11323 MW; BD8B396EE75F94FB CRC64;

Query Match 2.1%; Score 7; DB 6; Length 109;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 277 DFLFNGV 283
 |||||
 DB 27 DFLFNGV 33

RESULT 14

O9AF25 PRELIMINARY; PRT; 124 AA.
 AC O9AF25; (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DE Orf, hypothetical.
 GN S0014.
 OS Shigella flexneri.
 CC Plasmid virulence plasmid pWR501.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Shigella.
 OX NCBI_TaxID=623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-21189246; PubMed-11292750;
 RA Venkatesan M.M., Goldberg M.B., Rose D.J., Grobbeck E.J., Burland V.,
 RA Blattner F.R.;
 RT "Complete DNA sequence and analysis of the large virulence plasmid of
 RT Shigella flexneri";
 RL Infect. Immun. 69:3271-3285(2001).
 DR EMBL; AF348706; AAK18325.1; -;
 KW Plasmid.
 SQ SEQUENCE 124 AA; 13738 MW; 0A4B26442978BE1E CRC64;

Query Match 2.1%; Score 7; DB 2; Length 124;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 97 SKIRLLR 103
 |||||
 DB 91 SKIRLLR 97

RESULT 15

O8VSP8 PRELIMINARY; PRT; 124 AA.
 AC O8VSP8; (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DE Transposase-like protein.
 GN CP0011.
 OS Shigella flexneri 2a.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Shigella.
 OX NCBI_TaxID=42897;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-301;
 RA Jin O., Zhang J.Y., Liu H., Yang J., Yang F., Zhang X.B., Wang J.H.,
 RA Yang G.W., Wu H.T., Dong J., Sun L.L., Xie Y., Zhao A.L., Gao Y.S.,
 RA Zhu J.P., Kan B., Chen S.X., Yao Z.J., He B.K., Chen R.S., Ma D.L.,
 RA Yuan Z.H., Xu J.G., Wang Y., Shen Y., Lu W.C., Qiang B.Q., Wen Y.M.,
 RA Hou Y.D.;
 RT "Complete DNA sequence and analysis of the large virulence plasmid
 RT pCP301 of Shigella flexneri";
 RL Submitted (May-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF386526; AAL72476.1; -;
 KW Plasmid.
 SQ SEQUENCE 124 AA; 13765 MW; 1BBB37A5D978BE03 CRC64;

Query Match 2.1%; Score 7; DB 2; Length 124;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 97 SKIRLLR 103
 |||||
 DB 91 SKIRLLR 97

Wed Jun 25 10:49:28 2003

us-09-870-406a-33.oli.rspt

Page 6

Search completed: June 24, 2003, 18:55:43
Job time : 29 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 18:31:45 ; Search time 44 Seconds

(without alignments)
1540.670 Million cell updates/sec

Title: US-09-870-406a-33

Perfect score: 1724

Sequence: 1 MALSMERFGSICGSKAPNP.....SIIGKSYETRPMDITGQFIG 329

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SPREMBL_21:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_plant:*
11: sp_virus:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1724	100.0	329	10 Q9SR43	Q9SR43 arabidopsis
2	1400	81.2	273	10 Q9SR72	Q9SR72 arabidopsis
3	592	34.3	240	10 Q8R2B6	Q8R2B6 oryza sativ
4	339.5	19.7	255	2 Q93TM8	Q93TM8 nostoc punc
5	217.5	12.6	257	2 Q93WM9	Q93WM9 synecococc
6	215	12.5	257	2 Q9K4U5	Q9K4U5 prochloroco
7	205	11.9	280	2 Q93TL6	Q93TL6 nostoc punc
8	172.5	10.0	236	2 Q93SN7	Q93SN7 prochloroco
9	135	7.8	256	2 Q93SN8	Q93SN8 prochloroco
10	122.5	7.1	241	2 Q9K4U6	Q9K4U6 prochloroco
11	121.5	7.0	245	16 Q93TN0	Q93TN0 anabaena sp
12	109.5	6.4	245	2 Q93TM9	Q93TM9 nostoc punc
13	109.5	6.4	349	2 Q31193	Q31193 bacillus st
14	106.5	6.2	549	2 Q9K4U6	Q9K4U6 bacillus st
15	105.5	6.1	248	16 Q55891	Q55891 synecocyst
16	102.5	5.9	311	2 Q30422	Q30422 caldocellum

17	101.5	5.9	492	13 Q98796	Q98796 brachydanio
18	98	5.7	276	10 Q9H14	Q9H14 arabidopsis
19	97.5	5.7	521	11 Q8R3A8	Q8R3A8 mus musculus
20	96.5	5.6	241	2 Q93TL5	Q93TL5 prochloroco
21	96	5.6	852	4 Q96N04	Q96N04 homo sapien
22	95.5	5.5	595	16 Q92F27	Q92F27 rickettsia
23	95	5.5	574	2 Q99002	Q99002 shigella fl
24	95	5.5	1723	5 Q8WRD0	Q8WRD0 plasmodium
25	94.5	5.5	369	16 Q8RMA9	Q8RMA9 fusobacteri
26	94	5.5	392	10 Q8VX13	Q8VX13 arabidopsis
27	94	5.5	1452	16 Q97DNO	Q97DNO clostridium
28	93.5	5.4	386	16 Q9C1Y3	Q9C1Y3 lactococcus
29	93.5	5.4	1680	4 Q9P129	Q9P129 homo sapien
30	93.5	5.4	2819	16 Q980P8	Q980P8 mycoplasma
31	93	5.4	289	2 Q924M9	Q924M9 legionella
32	93	5.4	289	2 Q924N0	Q924N0 legionella
33	93	5.4	289	2 Q923C9	Q923C9 legionella
34	93	5.4	289	2 Q92374	Q92374 legionella
35	92.5	5.4	673	5 Q9TYX1	Q9TYX1 caenorhabdi
36	92	5.3	747	3 Q94058	Q94058 candida alb
37	92	5.3	1348	5 Q9VD13	Q9VD13 drosophila
38	91.5	5.3	841	17 Q59612	Q59612 pyrococcus
39	91.5	5.3	962	11 Q9JL18	Q9JL18 mus musculu
40	91.5	5.3	1382	4 Q9BX80	Q9BX80 homo sapien
41	91.5	5.3	3119	5 Q25857	Q25857 plasmodium
42	91	5.3	830	5 Q9VNZ7	Q9VNZ7 drosophila
43	91	5.3	1073	3 Q06839	Q06839 saccharomyc
44	91	5.3	2091	16 Q97J50	Q97J50 clostridium
45	91	5.3	2100	3 P87112	P87112 schizosacch

ALIGNMENTS

RESULT 1

ID	Q9SR43	PRELIMINARY:	PRT:	329 AA.
AC	Q9SR43:			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	F3124.1 protein (PHYTOCHROMOBILIN synthase HY2 protein).			
GN	F3124.1 OR HY2.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosids II; Brassicales; Brassicaceae; Arabidopsids.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV, COLUMBIA;			
RA	Lln X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,			
RA	Ronning C.M., Koo H., Fujii C.Y., Uteback T.R., Barnstead M.E.,			
RA	Bowman C.L., White O., Nierman W.C., Fraser C.M.;			
RT	"Arabidopsis thaliana chromosome III BAC F3124 genomic sequence."			
RL	Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21124703; PubMed=11226195;			
RA	Koichi T., Mukougawa K., Frankenberg N., Masuda M., Yokota A.,			
RA	Lagaras J.C.;			
RT	"The Arabidopsis HY2 Gene Encodes Phytochromobilin Synthase, a			
RT	Ferredoxin-Dependent Biliiverdin Reductase."			
RL	Plant Cell 13:425-436(2001).			
DR	EMBL; AC011436; AAF14017.1; -			
DR	EMBL; AB045112; BAB33374.1; -			
SO	SEQUENCE 329 AA; 38129 MW; 98727CD1778D973C CRC64;			
Query Match	100.0%;	Score 1724;	DB 10;	Length 329;
Best Local Similarity	100.0%;	Pred. No. 6.2e-136;		
Matches 329;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1 MALSMERFGSICGSKAPNPVLLSASPKNTNFTLRKRKKRFLRVSAVSYKEFAESALE 60			

```

Db 1 MALSMERFSGISGCKAPNPVLISASPNKINFTLRKKRKFLLKSAVSKKEPAESALE 60
QY 61 ETRKRIVLEPSSHLOKYSMTGLDQKTELQMLAFKSSKIRLLRSMALINETMVOFDPAGE 120
Db 61 ETRKRIVLEPSSHLOKYSMTGLDQKTELQMLAFKSSKIRLLRSMALINETMVOFDPAGE 120
QY 121 MEPEYDPIFCANFETSTNNIVVLDLNLPLHQLDQDQDYODKYKNTMSYHKYAEFFPW 180
Db 121 MEPEYDPIFCANFETSTNNIVVLDLNLPLHQLDQDQDYODKYKNTMSYHKYAEFFPW 180
QY 181 GSKLTGESIKFSPFLVMTWTRSSSKKHKALFSAFLEYQAWLEMTIOVREMEPSHYRA 240
Db 181 GSKLTGESIKFSPFLVMTWTRSSSKKHKALFSAFLEYQAWLEMTIOVREMEPSHYRA 240
QY 241 NCEAOKHYLTWRAQKDPGHGLKRLVGEAKAKELLDLFLNGVDELGTFTIDYFPEYOT 300
Db 241 NCEAOKHYLTWRAQKDPGHGLKRLVGEAKAKELLDLFLNGVDELGTFTIDYFPEYOT 300
QY 301 EDGTVSDKRSIIIGKSYETRPMDLTGQFIG 329
Db 301 EDGTVSDKRSIIIGKSYETRPMDLTGQFIG 329

```

RESULT 2

```

ID Q9S572 PRELIMINARY; PRT; 273 AA.
AC Q9S572;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
DE MZB10.18 protein.
GN MZB10.18.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RC SEQUENCE FROM N.A.
RA STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Rensing C.M., Koo H., Fujii C.Y., Utterback T.R., Bannstead M.E.,
RA Bowman C.L., White O., Nleman W.C., Fraser C.M.;
RA "Arabidopsis thaliana chromosome III p1 MZB10 genomic sequence.";
RA Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL: AC009326; AAD56331.1;
DR EMBL: AC009326; AAD56331.1;
SQ SEQUENCE 273 AA; 31909 MW; 06E942276466E87 CRC64;

```

Query Match 81.2%; Score 1400; DB 10; Length 273;
 Best Local Similarity 83.0%; Pred. No. 5.5e-109;
 Matches 273; Conservative 0; Mismatches 0; Indels 56; Gaps 2;

```

QY 1 MALSMERFSGISGCKAPNPVLISASPNKINFTLRKKRKFLLKSAVSKKEPAESALE 60
Db 1 MALSMERFSGISGCKAPNPVLISASPNKINFTLRKKRKFLLKSAVSKKEPAESALE 60
QY 61 ETRKRIVLEPSSHLOKYSMTGLDQKTELQMLAFKSSKIRLLRSMALINETMVOFDPAGE 120
Db 61 ETRKRIVLEPSSHLOKYSMTGLDQKTELQMLAFKSSKIRLLRSMALINETMVOFDPAGE 120
QY 121 MEPEYDPIFCANFETSTNNIVVLDLNLPLHQLDQDQDYODKYKNTMSYHKYAEFFPW 180
Db 121 MEPEYDPIFCANFETSTNNIVVLDLNLPLHQLDQDQDYODKYKNTMSYHKYAEFFPW 180
QY 181 GSKLTGESIKFSPFLVMTWTRSSSKKHKALFSAFLEYQAWLEMTIOVREMEPSHYRA 240
Db 181 GSKLTGESIKFSPFLVMTWTRSSSKKHKALFSAFLEYQAWLEMTIOVREMEPSHYRA 240
QY 241 NCEAOKHYLTWRAQKDPGHGLKRLVGEAKAKELLDLFLNGVDELGTFTIDYFPEYOT 300
Db 241 NCEAOKHYLTWRAQKDPGHGLKRLVGEAKAKELLDLFLNGVDELGTFTIDYFPEYOT 300

```

```

QY 301 EDGTVSDKRSIIIGKSYETRPMDLTGQFIG 329
Db 245 EDGTVSDKRSIIIGKSYETRPMDLTGQFIG 273

```

RESULT 3

```

ID Q9R2B6 PRELIMINARY; PRT; 240 AA.
AC Q9R2B6;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, last sequence update)
DE B147A04.9 protein.
GN B147A04.9.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Euphorbiaceae; Oryzae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RC SEQUENCE FROM N.A.
RA STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RA "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
RT clone:B147A04.9";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF003735; BAB86192.1;
SQ SEQUENCE 240 AA; 27459 MW; 7C3433A045FE9E2 CRC64;

```

Query Match 34.3%; Score 592; DB 10; Length 240;
 Best Local Similarity 44.3%; Pred. No. 1.4e-41;
 Matches 124; Conservative 29; Mismatches 65; Indels 62; Gaps 3;

```

QY 51 YKFPASALETRKRIVLEPSSHLOKYSMTGLDQKTELQMLAFKSSKIRLLRSMALINETMVOFDPAGE 110
Db 17 YKFPASALETRKRIVLEPSSHLOKYSMTGLDQKTELQMLAFKSSKIRLLRSMALINETMVOFDPAGE 110
QY 111 -TMQVDFAGFMEPEYDPIFCANFETSTNNIVVLDLNLPLHQLDQDQDYODKYKNTMSYHKYAEFFPW 169
Db 77 NSQVLDLRAFPSEPTDPIFCANFETSTNNIVVLDLNLPLHQLDQDQDYODKYKNTMSYHKYAEFFPW 169
QY 170 YKFAETFFPWGKLTGESIKFSPFLVMTWTRSSSKKHKALFSAFLEYQAWLEMTIOVREMEPSHYRA 229
Db 137 LVHKYNE-----LVMEFLDMDA 152
QY 230 REEMPSHYRANCEAOKHYLTWRAQKDPGHGLKRLVGEAKAKELLDLFLNGVDELGTFTIDYFPEYOT 289
Db 153 IKENKATIANNOEORHKYLTWRAEKD-----LVMEFLDMDA 152
QY 290 TFIDYFPEYOTEDGTVSDKRSIIIGKSYETRPMDLTGQFIG 329
Db 196 SFLDYFPEYARDDGSVKNKRSIMGKSFETRPMDANGFEFG 235

```

RESULT 4

```

ID Q93TM6 PRELIMINARY; PRT; 255 AA.
AC Q93TM6;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
DE B147A04.9 protein.
GN B147A04.9.
OS Nostoc punctiforme.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=63737;
RN [1]
RC SEQUENCE FROM N.A.
RA MEDLINE=21178821; PubMed=11283349;
RA Fraenkenberg N., Mukougawa K., Kohchi T., Lagarias J.C.;
RA "Functional genomic analysis of the HY2 family of ferredoxin-dependent
RT bilin reductases from oxygenic photosynthetic organisms.";
RL Plant Cell 13:965-978(2001).

```



```

Db      1 MLNQSQSLRNVALINMETCMATITVFARFNKSCSMKPFLELEKELFORPDIQSRVI 60
QY      70 PSHLOEKYSMTGIDGKTELQMLAFKSSKIRLRSMAIE-NEIMOVDFPAFMPEYDTP 128
Db      61 PGGLEFVSD-RGNRPAT-INSWCYQSOELKIRITYIDAGESAQIFNSVYPSHNDLP 118
QY      129 IFCANFTSTNV-NIVVLDLNPLOLDQDQDYQDKYKIMSIYHKAETFPWGGKLTGE 187
Db      119 LAGIDFLSGKVKMLIYLDFOPLFQ---DEDYQNKYIAPLKYLNKYPDLAQNEMKYD 175
QY      188 SIKFSPVMTFRSSSKKKAIFSAFLEYQAMLEKTIQVREEMPSHVANCEAOKH 247
Db      176 AMQFESKYLLEAK-TDEETVSTRVEAFQDYLDLWQMLADADALHDPEDIORIVKAKD 234
QY      248 YLTRAQDPGPHGLKRLVGEAKAKELLRDLF 280
Db      235 YDOYSADRDPAASGLFSSYFGHEMAERFLHERLF 267

```

RESULT 8

Q93SN7 PRELIMINARY; PRT; 257 AA.

```

ID      093SN7
AC      093SN7
DT      01-DEC-2001 (TREMblrel. 19, Created)
DT      01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT      01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE      Putative phycoerythrobilin:ferredoxin oxidoreductase.
GN      PEBB.
OS      Prochlorococcus sp. (strain CCMP 1378 / MED4).
OC      Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococaceae;
OC      Prochlorococcus.
OX      NCBI_Taxid=59919;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=MED4;
RX      MEDLINE=21178821; PubMed=11283349;
RA      Frankenberg N., Mukougawa K., Koichi T., Lagarias J.C.;
RT      "Functional genomic analysis of the H2 family of ferredoxin-dependent
RL      bilin reductases from oxygenic photosynthetic organisms.";
DR      EMBL: AY030301; AAK38142.1; -.
SQ      SEQUENCE 257 AA; 29872 MW; 02960652689AD58 CRC64;

```

Query Match 10.0%; Score 172.5; DB 2; Length 257;
 Best Local Similarity 21.2%; Pred. No. 1.6e-06;
 Matches 55; Conservative 52; Mismatches 134; Indels 19; Gaps 5;

```

QY      29 NKINFTLRRRRKRLVSAVSYPEFASALE---TRKRIVLEPSHLOEKYSMTGLD 85
Db      5 NTIFISOEWMAKFT-----KFLISQLDNYHCVEHKTASDFSYSKSSYSGSKSKKN 55
QY      86 KTELQMLAFKSKIRLRSMAIENETQVDFAGFMEPEYDTPIFCANFTSTNVNIVL 145
Db      56 INFTWGATGOKRINFRACVINSVNLFLIPTSTINIPFLGADVSLPSHLVL 115
QY      146 DLNPLHQLDQDQDYQDKYKIMSIYHKAETFPWGGKLTGESIKFSPVMTFRSSSK 205
Db      116 DFOF--SLKVENQFNSLEIQLIKKSCSSLPVAKMSVEDAKFFSPGLISRLAKHQ 173
QY      206 EK----HKAIFSAFLEYQAMLEKTIQVREEMPSHVANCEAOKHLYTRAQDPGPHGL 261
Db      174 DSNLNIENQLYDSFKELINLYLK-TLESEYVGHGLQDELINGQNDLYNRNDPAPRM 232
QY      262 LKRLVGEAKAKELLRDLFN 281
Db      233 LSSLPGKDPTESLINKVLF 252

```

RESULT 9
 Q93SN8 PRELIMINARY; PRT; 236 AA.
 AC Q93SN8;

```

DT      01-DEC-2001 (TREMblrel. 19, Created)
DT      01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT      01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE      Putative 15,16-dihydrobiliverdin:ferredoxin oxidoreductase.
GN      PEBB.
OS      Prochlorococcus sp. (strain CCMP 1378 / MED4).
OC      Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococaceae;
OC      Prochlorococcus.
OX      NCBI_Taxid=59919;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=MED4;
RX      MEDLINE=21178821; PubMed=11283349;
RA      Frankenberg N., Mukougawa K., Koichi T., Lagarias J.C.;
RT      "Functional genomic analysis of the H2 family of ferredoxin-dependent
RL      bilin reductases from oxygenic photosynthetic organisms.";
DR      EMBL: AY030300; AAK38141.1; -.
SQ      SEQUENCE 236 AA; 28368 MW; 783C086599C351D3 CRC64;

```

Query Match 7.8%; Score 135; DB 2; Length 236;
 Best Local Similarity 23.5%; Pred. No. 0.002;
 Matches 57; Conservative 53; Mismatches 103; Indels 30; Gaps 10;

```

QY      50 SYKFEASALEETRRKRIVLEPSHLO--EKYSMTGLDQKTELQMLAFKSKIRLRSMAI 107
Db      4 SLKRFVATNIED-----LDGKELIISKREKHNKSKYILKMWIFESQYRKWRITKL 57
QY      108 E-NETQVDFPAGMEPEYDTPIFCAN--FFTSTNVNIVVLDLNPLOLDQDQDYQDKY 164
Db      58 DGGDKLOVFTVAAPNPKSEPIIGADILWFTSQKLALFDYQPLIQ---EKYVLOKYC 114
QY      165 NKIMSIYHKAETFPWGGKLTGESIKFSPVMTFR--FSSSEKKAIFSAFLEYQOA 221
Db      115 SSDLFIKNQYSPDNHMKKNITVDSKTFSPWMLCRKNKLNLDRLNN-IFCSFVSNY-- 171
QY      222 WLENTI---QVREEMPSHVANCEAOKHLYTRAQDPGPHGLKRLVGEAKAKELLRD 277
Db      172 ---LTINKLHQNNQFLDLEQIKNN---QIDYDKYSAEKDPAADKLKTFPGFETWTFINN 225
QY      278 FLF 280
Db      226 FLF 228

```

RESULT 10

Q9K4U6 PRELIMINARY; PRT; 241 AA.

```

ID      09K4U6
AC      09K4U6;
DT      01-OCT-2000 (TREMblrel. 15, Created)
DT      01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT      01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE      Hypothetical 28.7 kDa protein.
OS      Prochlorococcus marinus.
OC      Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococaceae;
OC      Prochlorococcus.
OX      NCBI_Taxid=1219;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=SS120;
RA      Irlbacher H.M., Hess W.R.;
RT      "Organization of the ORF241/257 coding region from Prochlorococcus
RL      marinus SS120.";
DR      Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
EMBL: AJ278499; CAB95700.1; -.
KW      Hypothetical protein.
SQ      SEQUENCE 241 AA; 28699 MW; 4569F322DC35BBFC CRC64;

```

Query Match 7.1%; Score 122.5; DB 2; Length 241;
 Best Local Similarity 26.6%; Pred. No. 0.022;
 Matches 51; Conservative 29; Mismatches 79; Indels 33; Gaps 9;

QY 98 KIRLRSMAIENETQVDFPAGMEPEY--DPIFCAN--FFTSTNVNIVVLDLNPLOLD 153

```

Db      66 KLOVLSNVAY-----PAYTNDKPLIGDILMFGLKRLKRYAVLDEQPLVQ- 109
Oy      154 TDQDIDYQDKYNTKMSIYHKAETFPWCGKLTGESIKFSPVLMYMT--FSSSEKKHAL 211
Db      110 --EERFYCRKYKDIQILKNRVDNFNSOKTKRYISNKFSPVLLYNGSFDDLCSTLAKI 167
Oy      212 FSAFLEYQAWLEMTIQVREME--PSHYRANCEAQH-KYLTWRQKDPGHGLKRLAGE 268
Db      168 LDFELHAY--WQVNNNSREYIKIIPSKV----FQHTNYDIYSAERDPAHGLEKSYFGQ 221
Oy      269 AKAKELRLDFLE 280
Db      222 TWADQFVREFLE 233

RESULT 11
Oy3TN0 PRELIMINARY; PRT; 245 AA.
AC 093TN0:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Phycocyanobilin:ferredoxin oxidoreductase (Alr3707 protein).
GN PCYA OR ALR3707.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21178821; PubMed=11283349;
RA Frankenberg N., Mukougawa K., Kohchi T., Lagarias J.C.;
RT "Functional genomic analysis of the H2 family of ferredoxin-dependent
  bilin reductases from oxygenic photosynthetic organisms.";
RL Plant Cell 13:965-978(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneke T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
  Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
  Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
  Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
  Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
  cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AF339056; AAK38587.1; -
DR EMBL; AP003594; BAB75406.1; -
RW Complete proteome.
SQ SEQUENCE 245 AA; 27959 MW; 8F6CE652FEED011A CRC64;

Query Match      7.0%; Score 121.5; DB 16; Length 245;
Best Local Similarity 21.6%; Pred. No. 0.028;
Matches 52; Conservative 49; Mismatches 113; Indels 27; Gaps 10;

```

```

Db      242 D 242
RESULT 12
Oy3TM9 PRELIMINARY; PRT; 245 AA.
AC 093TM9:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Phycocyanobilin:ferredoxin oxidoreductase.
GN PCYA.
OS Nostoc punctiforme.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=63737;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21178821; PubMed=11283349;
RA Frankenberg N., Mukougawa K., Kohchi T., Lagarias J.C.;
RT "Functional genomic analysis of the H2 family of ferredoxin-dependent
  bilin reductases from oxygenic photosynthetic organisms.";
RL Plant Cell 13:965-978(2001).
DR EMBL; AF339057; AAK38588.1; -
SQ SEQUENCE 245 AA; 28061 MW; 20B9A8984F9C64C4 CRC64;

Query Match      6.4%; Score 109.5; DB 2; Length 245;
Best Local Similarity 20.9%; Pred. No. 0.28;
Matches 51; Conservative 50; Mismatches 110; Indels 33; Gaps 12;

```

```

Oy      52 KEFASALEETRRKRIVLEPSHLOEKYSMTG-LDG-KTELQMLAFKSKIRLLR-SMAIE 108
Db      18 RQLADCIENAHQHLDLSPYHLRDLGYEGRLSEKLTIERKCYQYQFRKMHLELANI 77
Oy      109 NETMQVEDFAGFMEPEYDPIF-CANFTSTNVNIVLDLPLHQLTDQTDYQDKYKNI 167
Db      78 GNMLDILHCVFMPREYDPIFPGCDLVGGRQISALIDLSPV--LDRT-LPEEYNAL 134
Oy      168 MSI-----YHKAETFPWCGKLTGESIKFSPVLMYMT--FSSSEKKHALFSA----FLEY 219
Db      135 AOLPLVNLSPQRELDWGC-----NIFSDCIFVRGSPSE--EAMFLSVREFLDIH 184
Oy      220 --QAWLEMTIQVREEMESHVRANCEAOHKYLTWRQKDPGHGLKRLVGEAKAKELLRD 277
Db      185 CMQALASHPVSEQ-----VTQNLAGGHNTCTKQKQNDKRRVLEKAFGPMVMEYMTT 238
Oy      278 FLEFN 281
Db      239 VLFED 242

RESULT 13
Oy3I193 PRELIMINARY; PRT; 549 AA.
AC 031193:
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Alpha amylase.
GN AMI.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
  Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-ATCC 31195;
RA da Silva A.C.R., Fernandes E., Pueyo M.T.;
RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF032864; AAB86961.1; -
DR HSSP; P06278; IV05.
DR InterPro; IPR000461; Alpha-amylase.
DR Pfam; PF00128; alpha-amylase; 1.
SQ SEQUENCE 549 AA; 62651 MW; 2CA6899EDAC4D262 CRC64;

```

Query Match 6.4%; Score 109.5; DB 2; Length 549;
 Best Local Similarity 19.5%; Pred. No. 0.82;
 Matches 69; Conservative 50; Mismatches 116; Indels 119; Gaps 17;

QY 50 SYKEPESALEETRRKRIYEPHSHLOEKYSMTGLDQKTELQMLAFKSSKIRLRSMATEN 109
 DB 229 NYDYLMAYADLDMDHPEVYTELKNWGMVNTNIDG-----FRLDVAKH 272
 QY 110 ETMOVF-DFAGMEPEYDPIFCA-----NEFTSTNVNIVLDLNPLOLTDQ 156
 DB 273 IKSEFFEDMLSYRSQGRKLFYGEVWSYDINKLHNITKNGTMSLFDA-PLH----- 326
 QY 157 TDYODKY-----NKIMSIYHKYAEF-----PWGK 183
 DB 327 ---NKFTYASKSGAGFDMRTMTNTLMKDPPLAVFVDNHDTEPGALQSWVDW--- 379
 QY 184 LTGESIKFSPPLVMWTRFSSSEKHKALFSAFLEYQAMLEMTIQVREMEPSHYRANCE 243
 DB 380 -----FKPLA-YAFILTRQEGYPCVF--YGDYITIPQYNIPSLKSIDPLLIARDY 428
 QY 244 A---OHKYL-----WRAQ---KDPGHGLKRL-----VGEAKAKELLRDLF 280
 DB 429 AVGQHLYLHSDIIGWTRREGVTEKPGSLAALITDGGSKMAYVKGHAGKVFYDLTG 488
 QY 281 NGVDELGTFTIDYFPEYOTEDGTVS-----DKRSITGKSYETRPMDLTGOFI 328
 DB 489 NRSDEVITINS--DGMGEFKVNGSVSWVPKRTVTSTIARITTPW--TGEFV 538

RESULT 14

Q9KWT6 PRELIMINARY; PRT; 549 AA.
 AC Q9KWT6; 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DE 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE Alpha-amylase (EC 3.2.1.1).
 OS Bacillus stearothermophilus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Geobacillus.
 OX NCBI_TaxID=1422;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-US100;
 RA Bejar S.;
 DR Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL; Y17557; CAB93517.1; -.
 DR HSBP; P06278; IYUS.
 DR InterPro; IPR000461; Alpha_amylase.
 DR Pfam; PF00128; alpha-amylase; 1.
 DR PRINTS; PR00110; ALPHAAMYLASE.
 KW Glycosylase; Hydrolase.
 SQ SEQUENCE 549 AA; 62582 MW; 8DA3E66DF9120BCE CRC64;

Query Match 6.2%; Score 106.5; DB 2; Length 549;
 Best Local Similarity 19.5%; Pred. No. 1.5;
 Matches 69; Conservative 50; Mismatches 116; Indels 119; Gaps 17;

QY 50 SYKEPESALEETRRKRIYEPHSHLOEKYSMTGLDQKTELQMLAFKSSKIRLRSMATEN 109
 DB 229 NYDYLMAYADLDMDHPEVYTELKNWGMVNTNIDG-----FRLDVAKH 272
 QY 110 ETMOVF-DFAGMEPEYDPIFCA-----NEFTSTNVNIVLDLNPLOLTDQ 156
 DB 273 IKSEFFEDMLSYRSQGRKLFYGEVWSYDINKLHNITKNGTMSLFDA-PLH----- 326
 QY 157 TDYODKY-----NKIMSIYHKYAEF-----PWGK 183
 DB 327 ---NKFTYASKSGAGFDMRTMTNTLMKDPPLAVFVDNHDTEPGALQSWVDW--- 379
 QY 184 LTGESIKFSPPLVMWTRFSSSEKHKALFSAFLEYQAMLEMTIQVREMEPSHYRANCE 243

DB 380 -----FKPLA-YAFILTRQEGYPCVF--YGDYIGIPQYNIPSLKSIDPLLIARDY 428

QY 244 A---OHKYL-----WRAQ---KDPGHGLKRL-----VGEAKAKELLRDLF 280
 DB 429 AVGQHLYLHSDIIGWTRREGVTEKPGSLAALITDGGSKMAYVKGHAGKVFYDLTG 488

QY 281 NGVDELGTFTIDYFPEYOTEDGTVS-----DKRSITGKSYETRPMDLTGOFI 328
 DB 489 NRSDEVITINS--DGMGEFKVNGSVSWVPKRTVTSTIARITTPW--TGEFV 538

RESULT 15

Q55891 PRELIMINARY; PRT; 248 AA.
 AC Q55891; 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE Hypothetical protein slr0116.
 GN SLR0116.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96127529; PubMed=8590279;
 RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
 RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 RA Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 DR EMBL; D64004; BAAL0653.1; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 248 AA; 28129 MW; BEB10CEBCE5E515B CRC64;

Query Match 6.1%; Score 105.5; DB 16; Length 248;
 Best Local Similarity 21.4%; Pred. No. 0.61;
 Matches 50; Conservative 49; Mismatches 110; Indels 25; Gaps 10;

QY 58 ALEETRRKRIYEPHSHLOEKYSMTGLDQKTELQMLAFKSSKIRLRSMATENMOY 114
 DB 27 AIAASWQSLPKFQYLPEDLGYVEGRLEKVIENRCYQTPQFRKHNIELAVAGGLDI 86
 QY 115 FDRGMEPEYDPIF-CAFFSTSTNVNIVLDLNPLOLTDQRO-DKYRKINS--- 169
 DB 87 LHCVMPEPEYDPIFGLFQCDIVAGPGVSAIADISP-----YQSDRLPAAYOKSLAELG 141
 QY 170 --IYHKAETPPMGSKLTGESIKFSPPLVMWTRFSSSEKHKALFSAFLEYQAMLEMTI 227
 DB 142 OPEEEORELPPMG-----EIFSEYCLFIRSNVTEBER-FVGQVVDLQIHCHOST 192
 QY 228 QVREMEPSHYRANCEAOKHLYLWRAOKDPGHGLKRLVGEAKAKELLRDLF 281
 DB 193 -VAEPLSPAQTLERHGOQIHCCOQKNDKTRVLEKAFGEAVMERMSOVLDD 245

Search completed: June 24, 2003, 18:38:58
 Job time : 47 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 18:55:46 ; Search time 21 Seconds
(without alignments)
1695.240 Million cell updates/sec

Title: US-09-870-406a-33

Perfect score: 329

Sequence: 1 MALSMERFSGSCFKAPNP.....SIIGKSYETRPMDLTGQFIG 329

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 417779 seqs, 108206813 residues

Word size : 0

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB pep:*
- 2: /cgn2_6/ptodata/2/pubppaa/PTCT_NEW_PUB pep:*
- 3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB pep:*
- 4: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB pep:*
- 5: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB pep:*
- 6: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB pep:*
- 7: /cgn2_6/ptodata/2/pubppaa/PTCTUS_PUBCOMB pep:*
- 8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB pep:*
- 9: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB pep:*
- 10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB pep:*
- 11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB pep:*
- 12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB pep:*
- 13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB pep:*
- 14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	329	100.0	329	9 US-10-159-901-33	Sequence 33, Appl
2	329	100.0	329	9 US-10-159-901-34	Sequence 34, Appl
3	329	100.0	329	9 US-10-159-901-55	Sequence 55, Appl
4	329	100.0	329	9 US-09-870-406a-33	Sequence 33, Appl
5	329	100.0	329	9 US-09-870-406a-34	Sequence 34, Appl
6	329	100.0	329	9 US-09-870-406a-55	Sequence 55, Appl
7	2.1	46	10	US-09-864-761-42825	Sequence 42825, A
8	2.1	66	9	US-10-124-903-18	Sequence 18, Appl
9	2.1	100	10	US-09-764-853-468	Sequence 468, App
10	2.1	258	9	US-10-156-055A-1	Sequence 1, Appl
11	2.1	341	9	US-09-738-626-5668	Sequence 5668, Ap
12	2.1	360	9	US-09-925-299-861	Sequence 861, Appl
13	2.1	360	10	US-10-106-698-5377	Sequence 5377, Ap
14	2.1	360	10	US-09-925-299-861	Sequence 861, App
15	2.1	387	9	US-09-815-242-11793	Sequence 11793, A
16	2.1	398	10	US-10-060-432-29	Sequence 29, Appl
17	2.1	398	10	US-09-905-173-29	Sequence 29, Appl
18	2.1	472	10	US-09-815-242-10824	Sequence 10824, A
19	2.1	473	9	US-10-269-353-39	Sequence 39, Appl

20	7	2.1	473	10	US-09-790-264-39	Sequence 39, Appl
21	7	2.1	478	10	US-09-815-242-11835	Sequence 11835, A
22	7	2.1	574	9	US-09-374-046A-62	Sequence 62, Appl
23	7	2.1	663	10	US-09-741-669-444	Sequence 444, App
24	7	2.1	830	9	US-10-211-962-31	Sequence 31, Appl
25	7	2.1	830	9	US-10-211-962-36	Sequence 36, Appl
26	7	2.1	830	9	US-10-211-962-37	Sequence 37, Appl
27	7	2.1	830	9	US-10-211-962-38	Sequence 38, Appl
28	7	2.1	940	9	US-09-811-088-11	Sequence 11, Appl
29	7	2.1	1661	9	US-09-842-758-42	Sequence 42, Appl
30	6	1.8	8	9	US-09-876-904A-349	Sequence 349, App
31	6	1.8	8	9	US-09-876-904A-367	Sequence 367, App
32	6	1.8	13	9	US-10-118-079-26	Sequence 26, Appl
33	6	1.8	18	10	US-09-864-761-41176	Sequence 41176, A
34	6	1.8	18	10	US-09-864-761-41204	Sequence 41204, A
35	6	1.8	19	10	US-09-864-761-46763	Sequence 46763, A
36	6	1.8	21	10	US-09-853-830-85	Sequence 85, Appl
37	6	1.8	30	10	US-09-864-761-35754	Sequence 35754, A
38	6	1.8	36	10	US-09-864-761-36721	Sequence 36721, A
39	6	1.8	36	10	US-09-864-761-38362	Sequence 38362, A
40	6	1.8	37	10	US-09-864-761-42368	Sequence 42368, A
41	6	1.8	46	9	US-10-050-704-298	Sequence 298, App
42	6	1.8	47	9	US-10-230-318-13	Sequence 13, Appl
43	6	1.8	47	9	US-10-230-335-11	Sequence 11, Appl
44	6	1.8	48	10	US-09-864-761-35521	Sequence 35521, A
45	6	1.8	51	10	US-09-864-761-42474	Sequence 42474, A

ALIGNMENTS

RESULT 1
US-10-159-901-33
Sequence 33, Application US/10159901
Publication No. US20030073235A1
GENERAL INFORMATION:
APPLICANT: LAGARRIS, JOHN
APPLICANT: KOICHI, TAKAYUKI
APPLICANT: FRANKENBERG, NICOLE
APPLICANT: GABETTA, GREGORY
APPLICANT: MONTGOMERY, BERONDA
TITLE OF INVENTION: LIGHT CONTROLLED GENE EXPRESSION UTILIZING HETEROLOGOUS PHYTOC
FILE REFERENCE: 407T-907731US
CURRENT APPLICATION NUMBER: US/10/159,901
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: 60/294,463
PRIOR FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 57
SOFTWARE: PatentIn version 3.0
SEQ ID NO 33
LENGTH: 329
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-10-159-901-33

Query Match 100.0%; Score 329; DB 9; Length 329;
Best Local Similarity 100.0%; Pred. No. 3.7e-309;
Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALSMERFSGSCFKAPNPVLLISAPNKINFTLRKKRFLRLRSVASYKFFASALE 60
DB 1 MALSMERFSGSCFKAPNPVLLISAPNKINFTLRKKRFLRLRSVASYKFFASALE 60
QY 61 ETRKRIVLEPSHLOEYSSMTGIDGTELOMLAFKSSKIRLRSMAIENETQVDFPAGE 120
DB 61 ETRKRIVLEPSHLOEYSSMTGIDGTELOMLAFKSSKIRLRSMAIENETQVDFPAGE 120
QY 121 MEPEYPTPCAFNEFTSTANNIVLNLPHOITDQDQDKYNNKIMSYHRYAETFPW 180
DB 121 MEPEYPTPCAFNEFTSTANNIVLNLPHOITDQDQDKYNNKIMSYHRYAETFPW 180
QY 181 GGRIGESIKFSPPLMTWTRFSSSKERKALFAFLEYQAWLEMTIOVREMEPSHVA 240
DB 181 GGRIGESIKFSPPLMTWTRFSSSKERKALFAFLEYQAWLEMTIOVREMEPSHVA 240

```

Db      181 GGLTGESIKFSPPLVMTFRSSSEKHKALFSAFLEYOAMLEMTIOVREMEPSHVA 240
QY      241 NCEAOKHYLWRAQNDPGHGLKRLVGEAKARELLRDLFNGVDELGTFTFDPEYOT 300
Db      241 NCEAOKHYLWRAQNDPGHGLKRLVGEAKARELLRDLFNGVDELGTFTFDPEYOT 300
QY      301 EDGTVDKRSIIIGKSYETRPMDLTGQFTG 329
Db      301 EDGTVDKRSIIIGKSYETRPMDLTGQFTG 329

RESULT 2
US-10-159-901-34
; Sequence 34, Application US/10159901
; Publication No. US20030073235A1
; GENERAL INFORMATION:
; APPLICANT: LAGARIAS, JOHN
; APPLICANT: KOICHI, TAKAYUKI
; APPLICANT: FRANKENBERG, NICOLE
; APPLICANT: MONTGOMERY, BERONDA
; TITLE OF INVENTION: LIGHT CONTROLLED GENE EXPRESSION UTILIZING HETEROLOGOUS PHYTOCHROME
; FILE REFERENCE: 407T-907731US
; CURRENT APPLICATION NUMBER: US/10/159,901
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: 60/294,463
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 329
; TYPE: PRN
; ORGANISM: Arabidopsis thaliana

Query Match
Best Local Similarity 100.0%; Score 329; DB 9; Length 329;
Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 MALSMERGFSGCFKAPNPVPLISASPKINFTLRKKRRLRLRSVAVSKEFAESALE 60
Db      1 MALSMERGFSGCFKAPNPVPLISASPKINFTLRKKRRLRLRSVAVSKEFAESALE 60
QY      61 ETRKRIYLEPSHLOEKYSMTGLDGTTELQMLAFKSKIRLLRSMAIENETMOVDFPAGF 120
Db      61 ETRKRIYLEPSHLOEKYSMTGLDGTTELQMLAFKSKIRLLRSMAIENETMOVDFPAGF 120
QY      121 MEPEYDPIPCANFTSTNVNIYVLDLNPLOLTDOTDYODKYNNKINSIYHKAETFPW 180
Db      121 MEPEYDPIPCANFTSTNVNIYVLDLNPLOLTDOTDYODKYNNKINSIYHKAETFPW 180
QY      181 GGLTGESIKFSPPLVMTFRSSSEKHKALFSAFLEYOAMLEMTIOVREMEPSHVA 180
Db      181 GGLTGESIKFSPPLVMTFRSSSEKHKALFSAFLEYOAMLEMTIOVREMEPSHVA 180
QY      241 NCEAOKHYLWRAQNDPGHGLKRLVGEAKARELLRDLFNGVDELGTFTFDPEYOT 240
Db      241 NCEAOKHYLWRAQNDPGHGLKRLVGEAKARELLRDLFNGVDELGTFTFDPEYOT 240
QY      301 EDGTVDKRSIIIGKSYETRPMDLTGQFTG 329
Db      301 EDGTVDKRSIIIGKSYETRPMDLTGQFTG 329

RESULT 3
US-10-159-901-55
; Sequence 55, Application US/10159901
; Publication No. US20030073235A1
; GENERAL INFORMATION:
; APPLICANT: LAGARIAS, JOHN
; APPLICANT: KOICHI, TAKAYUKI
; APPLICANT: FRANKENBERG, NICOLE
; APPLICANT: MONTGOMERY, BERONDA
; TITLE OF INVENTION: LIGHT CONTROLLED GENE EXPRESSION UTILIZING HETEROLOGOUS PHYTOCHROME
; FILE REFERENCE: 407T-907731US
; CURRENT APPLICATION NUMBER: US/10/159,901
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: 60/294,463
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 33
; LENGTH: 329
; TYPE: PRN
; ORGANISM: Arabidopsis thaliana

Query Match
Best Local Similarity 100.0%; Score 329; DB 9; Length 329;
Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 MALSMERGFSGCFKAPNPVPLISASPKINFTLRKKRRLRLRSVAVSKEFAESALE 60
Db      1 MALSMERGFSGCFKAPNPVPLISASPKINFTLRKKRRLRLRSVAVSKEFAESALE 60
QY      61 ETRKRIYLEPSHLOEKYSMTGLDGTTELQMLAFKSKIRLLRSMAIENETMOVDFPAGF 120
Db      61 ETRKRIYLEPSHLOEKYSMTGLDGTTELQMLAFKSKIRLLRSMAIENETMOVDFPAGF 120
QY      121 MEPEYDPIPCANFTSTNVNIYVLDLNPLOLTDOTDYODKYNNKINSIYHKAETFPW 180
Db      121 MEPEYDPIPCANFTSTNVNIYVLDLNPLOLTDOTDYODKYNNKINSIYHKAETFPW 180
QY      181 GGLTGESIKFSPPLVMTFRSSSEKHKALFSAFLEYOAMLEMTIOVREMEPSHVA 180
Db      181 GGLTGESIKFSPPLVMTFRSSSEKHKALFSAFLEYOAMLEMTIOVREMEPSHVA 180
QY      241 NCEAOKHYLWRAQNDPGHGLKRLVGEAKARELLRDLFNGVDELGTFTFDPEYOT 240
Db      241 NCEAOKHYLWRAQNDPGHGLKRLVGEAKARELLRDLFNGVDELGTFTFDPEYOT 240
QY      301 EDGTVDKRSIIIGKSYETRPMDLTGQFTG 329
Db      301 EDGTVDKRSIIIGKSYETRPMDLTGQFTG 329

```

```

; APPLICANT: MONTGOMERY, BERONDA
; TITLE OF INVENTION: LIGHT CONTROLLED GENE EXPRESSION UTILIZING HETEROLOGOUS PHYTOCHROME
; FILE REFERENCE: 407T-907731US
; CURRENT APPLICATION NUMBER: US/10/159,901
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: 60/294,463
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 55
; LENGTH: 329
; TYPE: PRN
; ORGANISM: Arabidopsis thaliana

US-10-159-901-55

Query Match
Best Local Similarity 100.0%; Score 329; DB 9; Length 329;
Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 MALSMERGFSGCFKAPNPVPLISASPKINFTLRKKRRLRLRSVAVSKEFAESALE 60
Db      1 MALSMERGFSGCFKAPNPVPLISASPKINFTLRKKRRLRLRSVAVSKEFAESALE 60
QY      61 ETRKRIYLEPSHLOEKYSMTGLDGTTELQMLAFKSKIRLLRSMAIENETMOVDFPAGF 120
Db      61 ETRKRIYLEPSHLOEKYSMTGLDGTTELQMLAFKSKIRLLRSMAIENETMOVDFPAGF 120
QY      121 MEPEYDPIPCANFTSTNVNIYVLDLNPLOLTDOTDYODKYNNKINSIYHKAETFPW 180
Db      121 MEPEYDPIPCANFTSTNVNIYVLDLNPLOLTDOTDYODKYNNKINSIYHKAETFPW 180
QY      181 GGLTGESIKFSPPLVMTFRSSSEKHKALFSAFLEYOAMLEMTIOVREMEPSHVA 180
Db      181 GGLTGESIKFSPPLVMTFRSSSEKHKALFSAFLEYOAMLEMTIOVREMEPSHVA 180
QY      241 NCEAOKHYLWRAQNDPGHGLKRLVGEAKARELLRDLFNGVDELGTFTFDPEYOT 240
Db      241 NCEAOKHYLWRAQNDPGHGLKRLVGEAKARELLRDLFNGVDELGTFTFDPEYOT 240
QY      301 EDGTVDKRSIIIGKSYETRPMDLTGQFTG 329
Db      301 EDGTVDKRSIIIGKSYETRPMDLTGQFTG 329

RESULT 4
US-09-870-406A-33
; Sequence 33, Application US/09870406A
; Publication No. US20030104379A1
; GENERAL INFORMATION:
; APPLICANT: LAGARIAS, JOHN
; APPLICANT: KOICHI, TAKAYUKI
; APPLICANT: FRANKENBERG, NICOLE
; APPLICANT: MONTGOMERY, BERONDA
; TITLE OF INVENTION: HY2 FAMILY OF BILIN REDUCTASES
; FILE REFERENCE: 407T-907720US
; CURRENT APPLICATION NUMBER: US/09/870,406A
; PRIOR FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: 60/271,758
; PRIOR FILING DATE: 2000-06-08
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 33
; LENGTH: 329
; TYPE: PRN
; ORGANISM: Arabidopsis thaliana

US-09-870-406A-33

Query Match
Best Local Similarity 100.0%; Score 329; DB 9; Length 329;
Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 MALSMERGFSGCFKAPNPVPLISASPKINFTLRKKRRLRLRSVAVSKEFAESALE 60
Db      1 MALSMERGFSGCFKAPNPVPLISASPKINFTLRKKRRLRLRSVAVSKEFAESALE 60
QY      61 ETRKRIYLEPSHLOEKYSMTGLDGTTELQMLAFKSKIRLLRSMAIENETMOVDFPAGF 120
Db      61 ETRKRIYLEPSHLOEKYSMTGLDGTTELQMLAFKSKIRLLRSMAIENETMOVDFPAGF 120
QY      121 MEPEYDPIPCANFTSTNVNIYVLDLNPLOLTDOTDYODKYNNKINSIYHKAETFPW 180
Db      121 MEPEYDPIPCANFTSTNVNIYVLDLNPLOLTDOTDYODKYNNKINSIYHKAETFPW 180
QY      181 GGLTGESIKFSPPLVMTFRSSSEKHKALFSAFLEYOAMLEMTIOVREMEPSHVA 180
Db      181 GGLTGESIKFSPPLVMTFRSSSEKHKALFSAFLEYOAMLEMTIOVREMEPSHVA 180
QY      241 NCEAOKHYLWRAQNDPGHGLKRLVGEAKARELLRDLFNGVDELGTFTFDPEYOT 240
Db      241 NCEAOKHYLWRAQNDPGHGLKRLVGEAKARELLRDLFNGVDELGTFTFDPEYOT 240
QY      301 EDGTVDKRSIIIGKSYETRPMDLTGQFTG 329
Db      301 EDGTVDKRSIIIGKSYETRPMDLTGQFTG 329

```

RESULT 5
US-09-870-406A-34
; Sequence 34, Application US/09870406A
; Publication No. US20030104379A1

Query Match	100.0%	Score 329;	DB 9;	Length 329;
Best Local Similarity	100.0%	Pred. No. 3.7e-309;		
Matches 329; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

Dd	241 NCEAOKHYLLWRAQMDPGHGLLRUGEAKEALLRDLFNGVDELGTFEIDYFPEYQT 3000
Qy	301 EDGTVSDDKRSIIIGSKSYETRPMDLTGQFIG 329
Dd	301 EDGTVSDDKRSIIIGSKSYETRPMDLTGQFIG 329

Query Match	100.0%	Score 329;	DB 9;	Length 329;
Best Local Similarity	100.0%	Pred. No. 3.7e-309;		
Matches 329; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

RESULT 7
US-09-864-761-42825
; Sequence 42825, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:

: APPLICANT: Rank, David R.
 : APPLICANT: Hanzel, David K.
 : APPLICANT: Chen, Wensheng
 : TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 42825
LENGTH: 46
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AF186190.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.6
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.99
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
US-09-864-761-42825

Query Match 2.1%; Score 7; DB 10; Length 46;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 238 VRANCEA 244
DB 32 VRANCEA 38

RESULT 8
US-10-124-903-18
Sequence 18, Application US/10124903
Publication No. US20030037346A1
GENERAL INFORMATION:
APPLICANT: Craig, Roger
APPLICANT: Savakis, Charalambos
TITLE OF INVENTION: PROTEIN PRODUCTION SYSTEM
FILE REFERENCE: 18747/1032
CURRENT APPLICATION NUMBER: US/10/124,903

CURRENT FILING DATE: 2002-04-18
PRIOR APPLICATION NUMBER: PCT/GB 00/04013
PRIOR FILING DATE: 2000-10-19
PRIOR APPLICATION NUMBER: GB 9924721.5
PRIOR FILING DATE: 1999-10-19
PRIOR APPLICATION NUMBER: US 60/165,508
PRIOR FILING DATE: 1999-11-15
NUMBER OF SEQ ID NOS: 19
SOFTWARE: PatentIn version 3.1
SEQ ID NO 18
LENGTH: 66
TYPE: PRT
ORGANISM: camel
US-10-124-903-18

Query Match 2.1%; Score 7; DB 9; Length 66;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 DPGHGL 262
DB 22 DPGHGL 28

RESULT 9
US-09-764-853-468
Sequence 468, Application US/09764853
Patent No. US20020090672A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PJ206
CURRENT APPLICATION NUMBER: US/09/764,853
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 939
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 468
LENGTH: 100
TYPE: PRT
ORGANISM: Homo sapiens
US-09-764-853-468

Query Match 2.1%; Score 7; DB 10; Length 100;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 PVLISAS 27
DB 8 PVLISAS 14

RESULT 10
US-10-156-055A-1
Sequence 1, Application US/10156055A
Publication No. US20030022336A1
GENERAL INFORMATION:
APPLICANT: KIKKOMAN CORPORATION
TITLE OF INVENTION: A SORBITOL DEHYDROGENASE GENE, A NOVEL RECOMBINANT DNA, AND
FILE REFERENCE: 4853.0091-00000
CURRENT APPLICATION NUMBER: US/10/156,055A
CURRENT FILING DATE: 2002-09-03
PRIOR APPLICATION NUMBER: JP 2001-159870
PRIOR FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 258
TYPE: PRT
ORGANISM: Pseudomonas sp. KS-E1806
US-10-156-055A-1

Query Match 2.1%; Score 7; DB 9; Length 258;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 KRLVGEA 269
DB 211 KRLVGEA 217

RESULT 11

US-09-738-626-5668
Sequence 5668, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OKAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738, 626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 5668
LENGTH: 341
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5668

Query Match 2.1%; Score 7; DB 9; Length 341;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 FNGVDEL 286
DB 305 FNGVDEL 311

RESULT 12

US-09-925-299-861
Sequence 861, Application US/09925299
Publication No. US20030040617A9
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925, 299
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: PatentIn ver. 2.0
SEQ ID NO 861
LENGTH: 360
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE

LOCATION: (53)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (360)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-861

Query Match 2.1%; Score 7; DB 9; Length 360;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 268 EAKAKEL 274
DB 183 EAKAKEL 189

RESULT 13

US-10-106-698-5377
Sequence 5377, Application US/10106698
Publication No. US20030109690A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypept
FILE REFERENCE: PA005P1
CURRENT APPLICATION NUMBER: US/10/106, 698
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/157,137
PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/163,280
PRIOR FILING DATE: 1999-11-03
NUMBER OF SEQ ID NOS: 8564
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 5377
LENGTH: 360
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (53)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (360)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-5377

Query Match 2.1%; Score 7; DB 9; Length 360;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 268 EAKAKEL 274
DB 183 EAKAKEL 189

RESULT 14

US-09-925-299-861
Sequence 861, Application US/09925299
Patent No. US20020055627A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925, 299
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: PatentIn ver. 2.0
SEQ ID NO 861

LENGTH: 360
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (53)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (360)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-861

Query Match 2.1%; Score 7; DB 10; Length 360;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 268 EAKAKEL 274
DB 183 EAKAKEL 189

RESULT 15
US-09-815-242-11793
Sequence 11793, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11793
LENGTH: 387
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11793

Query Match 2.1%; Score 7; DB 10; Length 387;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 KEFAESA 58
DB 261 KEFAESA 267

Search completed: June 24, 2003, 19:00:23
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 10:44:09 ; Search time 3712 Seconds
(without alignments)
16934.822 Million cell updates/sec

Title: US-09-870-406a-32

Perfect score: 2160
Sequence: 1 gaattcccccagcagcagctg.....tcgtctctactcaaca 2160

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_ov:*
7: gb_ov:*
8: gb_ov:*
9: gb_ov:*
10: gb_ov:*
11: gb_ov:*
12: gb_ov:*
13: gb_ov:*
14: gb_ov:*
15: gb_ov:*
16: gb_ov:*
17: gb_ov:*
18: gb_ov:*
19: gb_ov:*
20: gb_ov:*
21: gb_ov:*
22: gb_ov:*
23: gb_ov:*
24: gb_ov:*
25: gb_ov:*
26: gb_ov:*
27: gb_ov:*
28: gb_ov:*
29: gb_ov:*
30: gb_ov:*
31: gb_ov:*
32: gb_ov:*
33: gb_ov:*
34: gb_ov:*
35: gb_ov:*
36: gb_ov:*
37: gb_ov:*
38: gb_ov:*
39: gb_ov:*
40: gb_ov:*
41: gb_ov:*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2160	100.0	85561	8	ATAC009326	AC009326 Arabidops
2	2160	100.0	106688	8	ATAC011436	AC011436 Arabidops
3	608	28.1	1358	8	AY099706	AY099706 Arabidops
4	567.2	26.3	1337	8	AB045112	AB045112 Arabidops
5	525.2	24.3	1308	8	AY088874	AY088874 Arabidops
6	343	15.9	1092	8	AY128900	AY128900 Arabidops
7	92.6	4.3	85561	8	ATAC009326	AC009326 Arabidops
8	91.6	4.2	150053	8	AP003735	AP003735 Oryza sat
9	53.6	2.5	7218	6	166454	166454 Sequence 14
10	49	2.3	12142	6	AX346574	AX346574 Sequence
11	48.4	2.2	1141	6	AX083744	AX083744 Sequence
12	48	2.2	188357	9	AL159996	AL159996 Human DNA
13	47.8	2.2	204782	2	AC011168	AC011168 Homo sapi
14	47	2.2	349980	6	AX344555	AX344555 Sequence
15	46.8	2.2	7441	6	AX348444	AX348444 Sequence
16	46.8	2.2	7479	6	AX339172	AX339172 Sequence
17	46.8	2.2	256774	2	AC116964	AC116964 Dictyoste
18	46.6	2.2	408	6	AX300942	AX300942 Sequence
19	46.6	2.2	54345	3	AC084152	AC084152 Caenorhab
20	46.4	2.1	15421	3	PF000174	PF000174 Sequence
21	46.4	2.1	17211	6	AX345557	AX345557 Sequence
22	46.2	2.1	15224	6	AX251963	AX251963 Sequence
23	46.2	2.1	15224	6	AX344355	AX344355 Sequence
24	46.2	2.1	15224	6	AX348750	AX348750 Sequence
25	45.6	2.1	23046	8	LES272306	LES272306 Lycopersi
26	45.4	2.1	11441	6	AX083744	AX083744 Sequence
27	45.4	2.1	7631	6	AX345763	AX345763 Sequence
28	45.4	2.1	17721	6	AX346630	AX346630 Sequence
29	45.4	2.1	18434	6	AX346909	AX346909 Sequence
30	45.4	2.1	103237	2	AC110484	AC110484 Rattus no
31	45.4	2.1	151120	2	AC015814	AC015814 Homo sapi
32	45.4	2.1	175001	2	AC066615	AC066615 Homo sapi
33	45.4	2.1	190776	2	AC064813	AC064813 Homo sapi
34	45.2	2.1	153421	2	CNS079P4	CNS079P4 Oryza sat
35	45	2.1	53150	2	AC116984	AC116984 Dictyoste
36	45	2.1	116252	9	AL445228	AL445228 Human DNA
37	44.8	2.1	11670	6	AX281284	AX281284 Sequence
38	44.8	2.1	67056	2	AC116919	AC116919 Dictyoste
39	44.8	2.1	268147	2	AC116966	AC116966 Dictyoste
40	44.6	2.1	5718	6	AX251218	AX251218 Sequence
41	44.6	2.1	5718	6	AX346275	AX346275 Sequence
42	44.6	2.1	171187	2	AC116960	AC116960 Dictyoste
43	44.4	2.1	10006	6	AX344938	AX344938 Sequence
44	44.4	2.1	183054	2	AC116074	AC116074 Rattus no
45	44.2	2.0	5216	6	AX281348	AX281348 Sequence

ALIGNMENTS

RESULT 1
ATAC009326 85561 bp DNA linear PLN 24-JAN-2001
LOCUS Arabidopsis thaliana chromosome III Pl M2B10 genomic sequence,
DEFINITION complete sequence.
ACCESSION AC009326
VERSION AC009326.8 GI:12408713
KEYWORDS HTG.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidops.
REFERENCE 1 (bases 1 to 85561)
AUTHORS Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.,

Rowing, C.M., Koo, H., Fujii, C.Y., Utterback, T.R., Barnstead, M.E.,
 Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.
 Arabidopsis thaliana chromosome III p1 MZB10 genomic sequence
 unpublished
 2 (bases 1 to 85561)
 Lin, X. and Kaul, S.
 Direct Submission
 Submitted (16-AUG-1999) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
 3 (bases 1 to 85561)
 Lin, X.
 Direct Submission
 Submitted (24-JUN-2001) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 On Jan 24, 2001 this sequence version replaced gi:12280842.
 Address all correspondence to:
 Xiaoyang Lin
 The Institute for Genomic Research
 9712 Medical Center Dr.
 Rockville, MD 20850, USA
 e-mail: xlin@tigr.org.
 Pl clone MZB10 is from Arabidopsis chromosome III and is near the
 molecular marker C1C7A12.
 The orientation of the sequence is from SP6 to T7 end of the pl
 clone.
 Genes were identified by a combination of three methods: Gene
 prediction programs including GRLIL (available by anonymous ftp
 from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
 Washington), GenScan (Chris Burge,
<http://genome.stanford.edu/~chris/GENSCANW.html>), and NetPlantgene
 (<http://www.cbs.dtu.dk/netplantgene/cbsnetplantgene.html>), searches of the
 complete sequence against a peptide database and the Arabidopsis
 EST database at TIGR (<http://www.tigr.org/tdb/at/est.html>).
 Annotated genes are named to indicate the level of evidence for
 their annotation. Genes with similarity to other proteins are named
 after the database hits. Genes without significant peptide
 similarity but with EST similarity are named as 'unknown' proteins.
 Genes without protein or EST similarity, that are predicted by more
 than two gene prediction programs over most of their length are
 annotated as 'hypothetical' proteins. Genes encoding tRNAs are
 predicted by tRNAscan-SE (Sean Eddy,
<http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are
 identified by repeatmasker (Arian Smit,
<http://ftp.genome.washington.edu/RM/repeatmasker.html>). Regions of
 genomic sequence that are not annotated as genes but have predicted
 exons by GRLIL are annotated as msc features.

FEATURES

Source

misc_feature	349. .702	/note="exon predicted by xgrall, quality excellent"
misc_feature	/note="exon predicted by xgrall, quality excellent"	complement(587. .681)
misc_feature	/note="exon predicted by xgrall, quality excellent"	shadowexon"
misc_feature	881. .1171	/note="exon predicted by xgrall, quality good"
misc_feature	1508. .1606	/note="exon predicted by xgrall, quality marginal"
misc_feature	complement(1516. .1602)	/note="exon predicted by xgrall, quality marginal"
misc_feature	1655. .1758	/note="exon predicted by xgrall, quality marginal"
misc_feature	1655. .1758	/note="exon predicted by xgrall, quality marginal"
misc_feature	complement(1682. .1733)	/note="exon predicted by xgrall, quality marginal"
misc_feature	complement(1852. .1922)	/note="exon predicted by xgrall, quality good, shadowexon"

	misc_feature	1854..1933 /note="exon predicted by xgrall, quality marginal" /db_xref="x281"
	gene	/gene="MZB10.1" /note="similar to MITOCHONDRIAL INNER MEMBRANE PROTEASE SUBUNIT 2 GB:P46972 [Saccharomyces cerevisiae]" join(<2185..2338,2432..2504,2922..3081,3204..>3281) /gene="MZB10.1" join(2185..2338,2432..2504,2922..3081,3204..3281) /gene="MZB10.1" /codon_start=1 /product="putative signal peptidase" /protein_id="AAD56314.1" /db_xref="GI:59236634" /translation="MGIONIMOVAKSKFTSGITGLTISDRCCSVVPVAGDSMPTFTF PQRNSYLDIVYDKRFLKDYKFRARGDVVFSSPTRFGRIYIKRIVGNGEVISSRR VIRPEGVGWEVSQGNKTSLSIDRSFGEIPLEGILGHVTREVMPPORISRIGH" complement(3284..3360) /note="exon predicted by xgrall, quality marginal shadowexon" complement(3372..3486) /note="exon predicted by xgrall, quality marginal shadowexon" complement(3872..3962) /note="exon predicted by xgrall, quality marginal shadowexon" complement(3922..4055) /rpt_family="(GAAA)n" <4496..>5463 /gene="MZB10.2" /note="similar to yippee putative zinc binding protein GB:AAD7882 [Homo sapiens]" join(<4496..4591,5070..5113,5199..5307,5386..>5463) /gene="MZB10.2" join(4496..4591,5070..5113,5199..5307,5386..5463) /gene="MZB10.2" /codon_start=1 /product="yippee-like protein" /protein_id="AAD56315.1" /db_xref="GI:59236644" /translation="MGRLEFVIDLEGIVSCRYCOTHPAATNDIISKSFCHKRGATLF DNVAVTVEKEHRWITCMHWADIFCVCSGLVGMKYEIAIDKSORYEKGFILER ESSEFN" complement(5490..5586) /note="exon predicted by xgrall, quality marginal shadowexon" 5534..5615 /note="exon predicted by xgrall, quality excellent" complement(5641..5725) /note="exon predicted by xgrall, quality excellent shadowexon" complement(6084..6095) /note="exon predicted by xgrall, quality marginal shadowexon" 6393..6476 /note="exon predicted by xgrall, quality excellent" complement(6418..6474) /note="exon predicted by xgrall, quality excellent shadowexon" complement(6632..6705) /rpt_family="(GA)n" <6920..>9232 /gene="MZB10.3" /note="similar to unknown protein GB:AAB95292 [Arabidopsis thaliana]" join(<6920..7010,7097..7253,7541..7646,7762..8769, 8969..>9232) /gene="MZB10.3" join(6920..7010,7097..7253,7541..7646,7762..8769, 8969..9232) /gene="MZB10.3" /note="hypothetical protein" /codon_start=1
CDS	misc_feature	1854..1933 /note="exon predicted by xgrall, quality marginal" /db_xref="x281"
	gene	/gene="MZB10.1" /note="similar to MITOCHONDRIAL INNER MEMBRANE PROTEASE SUBUNIT 2 GB:P46972 [Saccharomyces cerevisiae]" join(<2185..2338,2432..2504,2922..3081,3204..>3281) /gene="MZB10.1" join(2185..2338,2432..2504,2922..3081,3204..3281) /gene="MZB10.1" /codon_start=1 /product="putative signal peptidase" /protein_id="AAD56314.1" /db_xref="GI:59236634" /translation="MGIONIMOVAKSKFTSGITGLTISDRCCSVVPVAGDSMPTFTF PQRNSYLDIVYDKRFLKDYKFRARGDVVFSSPTRFGRIYIKRIVGNGEVISSRR VIRPEGVGWEVSQGNKTSLSIDRSFGEIPLEGILGHVTREVMPPORISRIGH" complement(3284..3360) /note="exon predicted by xgrall, quality marginal shadowexon" complement(3372..3486) /note="exon predicted by xgrall, quality marginal shadowexon" complement(3872..3962) /note="exon predicted by xgrall, quality marginal shadowexon" complement(3922..4055) /rpt_family="(GAAA)n" <4496..>5463 /gene="MZB10.2" /note="similar to yippee putative zinc binding protein GB:AAD7882 [Homo sapiens]" join(<4496..4591,5070..5113,5199..5307,5386..>5463) /gene="MZB10.2" join(4496..4591,5070..5113,5199..5307,5386..5463) /gene="MZB10.2" /codon_start=1 /product="yippee-like protein" /protein_id="AAD56315.1" /db_xref="GI:59236644" /translation="MGRLEFVIDLEGIVSCRYCOTHPAATNDIISKSFCHKRGATLF DNVAVTVEKEHRWITCMHWADIFCVCSGLVGMKYEIAIDKSORYEKGFILER ESSEFN" complement(5490..5586) /note="exon predicted by xgrall, quality marginal shadowexon" 5534..5615 /note="exon predicted by xgrall, quality excellent" complement(5641..5725) /note="exon predicted by xgrall, quality excellent shadowexon" complement(6084..6095) /note="exon predicted by xgrall, quality marginal shadowexon" 6393..6476 /note="exon predicted by xgrall, quality excellent" complement(6418..6474) /note="exon predicted by xgrall, quality excellent shadowexon" complement(6632..6705) /rpt_family="(GA)n" <6920..>9232 /gene="MZB10.3" /note="similar to unknown protein GB:AAB95292 [Arabidopsis thaliana]" join(<6920..7010,7097..7253,7541..7646,7762..8769, 8969..>9232) /gene="MZB10.3" join(6920..7010,7097..7253,7541..7646,7762..8769, 8969..9232) /gene="MZB10.3" /note="hypothetical protein" /codon_start=1
mRNA	misc_feature	1854..1933 /note="exon predicted by xgrall, quality marginal" /db_xref="x281"
	gene	/gene="MZB10.1" /note="similar to MITOCHONDRIAL INNER MEMBRANE PROTEASE SUBUNIT 2 GB:P46972 [Saccharomyces cerevisiae]" join(<2185..2338,2432..2504,2922..3081,3204..>3281) /gene="MZB10.1" join(2185..2338,2432..2504,2922..3081,3204..3281) /gene="MZB10.1" /codon_start=1 /product="putative signal peptidase" /protein_id="AAD56314.1" /db_xref="GI:59236634" /translation="MGIONIMOVAKSKFTSGITGLTISDRCCSVVPVAGDSMPTFTF PQRNSYLDIVYDKRFLKDYKFRARGDVVFSSPTRFGRIYIKRIVGNGEVISSRR VIRPEGVGWEVSQGNKTSLSIDRSFGEIPLEGILGHVTREVMPPORISRIGH" complement(3284..3360) /note="exon predicted by xgrall, quality marginal shadowexon" complement(3372..3486) /note="exon predicted by xgrall, quality marginal shadowexon" complement(3872..3962) /note="exon predicted by xgrall, quality marginal shadowexon" complement(3922..4055) /rpt_family="(GAAA)n" <4496..>5463 /gene="MZB10.2" /note="similar to yippee putative zinc binding protein GB:AAD7882 [Homo sapiens]" join(<4496..4591,5070..5113,5199..5307,5386..>5463) /gene="MZB10.2" join(4496..4591,5070..5113,5199..5307,5386..5463) /gene="MZB10.2" /codon_start=1 /product="yippee-like protein" /protein_id="AAD56315.1" /db_xref="GI:59236644" /translation="MGRLEFVIDLEGIVSCRYCOTHPAATNDIISKSFCHKRGATLF DNVAVTVEKEHRWITCMHWADIFCVCSGLVGMKYEIAIDKSORYEKGFILER ESSEFN" complement(5490..5586) /note="exon predicted by xgrall, quality marginal shadowexon" 5534..5615 /note="exon predicted by xgrall, quality excellent" complement(5641..5725) /note="exon predicted by xgrall, quality excellent shadowexon" complement(6084..6095) /note="exon predicted by xgrall, quality marginal shadowexon" 6393..6476 /note="exon predicted by xgrall, quality excellent" complement(6418..6474) /note="exon predicted by xgrall, quality excellent shadowexon" complement(6632..6705) /rpt_family="(GA)n" <6920..>9232 /gene="MZB10.3" /note="similar to unknown protein GB:AAB95292 [Arabidopsis thaliana]" join(<6920..7010,7097..7253,7541..7646,7762..8769, 8969..>9232) /gene="MZB10.3" join(6920..7010,7097..7253,7541..7646,7762..8769, 8969..9232) /gene="MZB10.3" /note="hypothetical protein" /codon_start=1
CDS	misc_feature	1854..1933 /note="exon predicted by xgrall, quality marginal" /db_xref="x281"
	gene	/gene="MZB10.1" /note="similar to MITOCHONDRIAL INNER MEMBRANE PROTEASE SUBUNIT 2 GB:P46972 [Saccharomyces cerevisiae]" join(<2185..2338,2432..2504,2922..3081,3204..>3281) /gene="MZB10.1" join(2185..2338,2432..2504,2922..3081,3204..3281) /gene="MZB10.1" /codon_start=1 /product="putative signal peptidase" /protein_id="AAD56314.1" /db_xref="GI:59236634" /


```

/protein_id="A056316.1"
/db_xref="GI:5923665"
/translation="MLTHRDDEISLFLENRREKREKRAADSLTSGDSVIMATLTA
AAALGSVSTASSQRYPLRRTAEENLYENERSDMDLTPGTQPFKEHSRYM
NOHDAPNSRPTSLKSLGNCREDIVSGNNKRPQSSSSGSLRPSGSSRSRA
TPRRSTPTSTSRPYTTRASNRSSTSRATLTAARATTAARITTSGSAR
SAPTRSNRPSSASSKKPYSRATPRRSTPGPISVSGSRSSRSTPSYNS
KAPSRGSPSPTLNSSRPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPK
SRGSLERGGPTSGSGNAROSCSRSRAPLGNNGSLTGVGRKAKSNGSGCD
NLSPPAANGMVERVNMKRLGPPRLNENGRSGSSAFNLSGYNLSKSIDMA
IRHMDIRGMTGNLPLVTKVPASMSYSGSVSSSPSVATSSVSSSDVDNIN
ILDLGNEANDLILERSYASAPRNOYPRFTS"
/misc_feature
/feature="exon predicted by xgrail, quality
marginal_shadowexon
10828..10857
/feature="exon predicted by xgrail, quality marginal"
<11221..>12992
/feature="MZB10.4"
/feature="similar to receptor kinase GB:S70769 [Arabidopsis
thaliana]"
mRNA
join(<11221..11263,11488..11600,11690..11900,11982..12219,
12298..12448,12597..>12992)
/feature="MZB10.4"
join(11221..11263,11488..11600,11690..11900,11982..12219,
12298..12448,12597..12992)
/feature="MZB10.4"
/codon_start=1
/product="putative receptor ser/thr protein kinase"
/protein_id="A056317.1"
/db_xref="GI:5923666"
/translation="MCGNSDRIGOREAEICTNNVREYSNLSRATDSHPNRIGS
GGYVPEKYLKRDGTQYAAVKSLSNESKQGRREPLEINLSIHHPVLKLGICIG
NMILVEYLENNSLASVLGSSRSRYPLDMSKRAALCVASGLAEHVEHVAH
RDIKASNIILDSNLSKPIGDFGLAKLEPDVHTVSTVAGTGLAEVYALGQLTK
ADYRSFQILVEVYISGNSSTRAFGDEYAVLVEYVNLREERLREDELTEFPD
EYRFIYALFCTQAAOKRPNKOVEMLRKELINENALPEYRGVNGRNRH
GIGRVGGSSGSSSESTORDKSSAVPGSISNSISOSTELAPR"
<14213..>15448
/feature="MZB10.5"
/feature="similar to hypothetical protein GB:AAC27164"
Query Match 100.0%; Score 2160; DB 8; Length 85561;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2160; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAATCCCAACGTCACGTGCTGATTCACGTGGCGGAGTGGCCCTATATGTTG 60
Db 64324 GAATCCCAACGTCACGTGCTGATTCACGTGGCGGAGTGGCCCTATATGTTG 64383
QY 61 ACCATGACTGGAGCGATGTTGAATTCATGTCGTCGCAATGCGCTTGTCTCAGTGA 120
Db 64384 ACCATGACTGGAGCGATGTTGAATTCATGTCGTCGCAATGCGCTTGTCTCAGTGA 64443
QY 121 AACTGTGAATTTTATCTCTTTATAGATAAGAAATCTGCTTTTTCAGTTTTCAGTA 180
Db 64444 AACTGTGAATTTTATCTCTTTATAGATAAGAAATCTGCTTTTTCAGTTTTCAGTA 64503
QY 181 TGAAGAAGATTGAAGAGTGTCCGAGAAAGACCTTGGTTTCAGTTTGTAGTCT 240
Db 64504 TGAAGAAGATTGAAGAGTGTCCGAGAAAGACCTTGGTTTCAGTTTGTAGTCT 64563
QY 241 TGTGTAAAGCTTATCAATGAGCTTGGTTTCAATGGGTCAGCTTCAAGGAGAC 300
Db 64564 TGTGTAAAGCTTATCAATGAGCTTGGTTTCAATGGGTCAGCTTCAAGGAGAC 64623
QY 301 AAACCCACCTGTTCTATCTCTGCAAGCCCTAAATAAATCAATTTACGTTGAGAAGAG 360
Db 64624 AAACCCACCTGTTCTATCTCTGCAAGCCCTAAATAAATCAATTTACGTTGAGAAGAG 64683
QY 361 AAAGAAAAGATTCTTACTAGAGTCTGCTGTGCTATPAGGAATTCGACAGTCTGC 420
Db 64684 AAAGAAAAGATTCTTACTAGAGTCTGCTGTGCTATPAGGAATTCGACAGTCTGC 64743

```

```

QY 421 TTTGAAAAGAACCGAAGAAAGATGCTTCTTGAACCTTCACATCTCCAGATATGCAAT 480
Db 64744 TTTGAAAAGAAACCGAAGAAAGATGCTTCTTGAACCTTCACATCTCCAGATATGCAAT 64803
QY 481 TACATTCGTAGTAGTGAGGAGATATATTTCTCATTTGTCTGCTGTAATTTTG 540
Db 64804 TACATTCGTAGTAGTGAGGAGATATATTTCTCATTTGTCTGCTGTAATTTTG 64863
QY 541 GGTAAATGATTGAGTGTGTCATTAGAGAACCAACAAATACCTTACTGTATAGACTGC 600
Db 64864 GGTAAATGATTGAGTGTGTCATTAGAGAACCAACAAATACCTTACTGTATAGACTGC 64923
QY 601 TTATATATAGTAAAGTTTCAGATTTTGTCTTCTATATACGAACTGTTTCAGAAAAGTA 660
Db 64924 TTATATATAGTAAAGTTTCAGATTTTGTCTTCTATATACGAACTGTTTCAGAAAAGTA 64983
QY 661 TAGTAGCATGACAGACTAGATGTTAGAGACCACTTCAATGCTTGTATTAATGCTC 720
Db 64984 TAGTAGCATGACAGACTAGATGTTAGAGACCACTTCAATGCTTGTATTAATGCTC 65043
QY 721 AAAGATTAGACTCTTGAGGAGTATGCAATAGAGAAATGACAAATGCAAGTTTAACCTCA 780
Db 65044 AAAGATTAGACTCTTGAGGAGTATGCAATAGAGAAATGACAAATGCAAGTTTAACCTCA 65103
QY 781 GCAGTACAAACTGATGCTTATGATCCCATTTCTTACTTCAATTTGATGATGTTGTGA 840
Db 65104 GCAGTACAAACTGATGCTTATGATCCCATTTCTTACTTCAATTTGATGATGTTGTGA 65163
QY 841 TCTTCGCTTAGCTTTGACTTTTCCGGGTTTCAGAGACCTGAGATATGATCTCCATAT 900
Db 65164 TCTTCGCTTAGCTTTGACTTTTCCGGGTTTCAGAGACCTGAGATATGATCTCCATAT 65223
QY 901 TCTGTGCTAACCTTTTTCACATCTACCAACGTTAACTATGTAATGTAATGTAATCTCT 960
Db 65224 TCTGTGCTAACCTTTTTCACATCTACCAACGTTAACTATGTAATGTAATGTAATCTCT 65283
QY 961 AGTAACTGAGATTAACAGCTGCTGATGTCGCAACGATGATCAATATTTACTGAT 1020
Db 65284 AGTAACTGAGATTAACAGCTGCTGATGTCGCAACGATGATCAATATTTACTGAT 65343
QY 1021 GTTCTCTTTAGGAGACCTTAATCTTTCATCACTGATGACAGAGATATACCAAGA 1080
Db 65344 GTTCTCTTTAGGAGACCTTAATCTTTCATCACTGATGACAGAGATATACCAAGA 65403
QY 1081 CAAGTATTAATACAAAGATATGTCATATATACAAATATGCTGAGTACACAGAT 1140
Db 65404 CAAGTATTAATACAAAGATATGTCATATATACAAATATGCTGAGTACACAGAT 65463
QY 1141 ACACCAATTAATCAATGCAAGTAAACCTTAATGCTGAGTAAATGAGATGATCTGAG 1200
Db 65464 ACACCAATTAATCAATGCAAGTAAACCTTAATGCTGAGTAAATGAGATGATCTGAG 65523
QY 1201 ATTATTTTGCAGACTTTCCATGGGAGGAGAAATGCACTGATCCATTAAGCTTTTTC 1260
Db 65524 ATTATTTTGCAGACTTTCCATGGGAGGAGAAATGCACTGATCCATTAAGCTTTTTC 65583
QY 1261 TCGCCTTTGGTGGAGTGGAGTCTGCTTTCGTCGCAAGAAACATTAAGCTTTTGTTC 1320
Db 65584 TCGCCTTTGGTGGAGTGGAGTCTGCTTTCGTCGCAAGAAACATTAAGCTTTTGTTC 65643
QY 1321 TCTGCGTTTCTAGAGTACTATCAGTATATCTCAGGCGCCAAAGCTTAAGCTTTTATG 1380
Db 65644 TCTGCGTTTCTAGAGTACTATCAGTATATCTCAGGCGCCAAAGCTTAAGCTTTTATG 65703
QY 1381 GAAACTTTGACTGGAATCTATCATCTTCTCTCAAGGCACTGCTGAGATGCAATCC 1440
Db 65704 GAAACTTTGACTGGAATCTATCATCTTCTCTCAAGGCACTGCTGAGATGCAATCC 65763
QY 1441 AAGTGAAGGAGAGATGGAACCATCTCATGTGAAGCCCAATTTGAAAGCACACAAAGT 1500
Db 65764 AAGTGAAGGAGAGATGGAACCATCTCATGTGAAGCCCAATTTGAAAGCACACAAAGT 65823

```

QY 1501 ACCGACATGCGGACACAAAAGGATTCATTTCCCTTTGTGTAATTCAGTGTGA 1560
 DB 65824 ACCGACATGCGGACACAAAAGGATTCATTTCCCTTTGTGTAATTCAGTGTGA 65883
 QY 1561 ACAGACATGCTGTAATGTTTCAATGATGATTTGGTTCAGATCCTGGA 1620
 DB 65884 ACAGACATGCTGTAATGTTTCAATGATGATTTGGTTCAGATCCTGGA 65943
 QY 1621 CATGCTCTTAAAGATTAGTAGTACCAAGCAAGCAAGTATAAAGATTGATCC 1680
 DB 65944 CATGCTCTTAAAGATTAGTAGTACCAAGCAAGCAAGTATAAAGATTGATCC 66003
 QY 1681 CATTAGTCTCCCATTTATTAATGCTTGTGAAGATGTTGAAAATGATTGACAAATC 1740
 DB 66004 CATTAGTCTCCCATTTATTAATGCTTGTGAAGATGTTGAAAATGATTGACAAATC 66063
 QY 1741 AGAGCTCTAAGGATTTCTCTGTTCAATGGGGTGGATGATGATGAGCAAAAACATTTCA 1800
 DB 66064 AGAGCTCTAAGGATTTCTCTGTTCAATGGGGTGGATGATGATGAGCAAAAACATTTCA 66123
 QY 1801 TTGATTTACTTCCAGAGTACCAACAGAAAGATGAGTAAAGCAAGATATCA 1860
 DB 66124 TTGATTTACTTCCAGAGTACCAACAGAAAGATGAGTAAAGCAAGATATCA 66183
 QY 1861 TTGGAGTCTATGAAAGTCTGTCATGGGATTTAAGAGCAATTTATGGCTAACAT 1920
 DB 66184 TTGGAGTCTATGAAAGTCTGTCATGGGATTTAAGAGCAATTTATGGCTAACAT 66243
 QY 1921 GATATATGTGAACAAGTCAATTTCAAGATCATCAACAAGAGAGCTGATAGGGA 1980
 DB 66244 GATATATGTGAACAAGTCAATTTCAAGATCATCAACAAGAGAGCTGATAGGGA 66303
 QY 1981 AGTAGGAATTAAGAAAGCAGCATGAGAGCTCTCTCAGGCTATCTGCAATTCAGATGA 2040
 DB 66304 AGTAGGAATTAAGAAAGCAGCATGAGAGCTCTCTCAGGCTATCTGCAATTCAGATGA 66363
 QY 2041 TTGTTGATTTACCATCATGATGATTTTACAAAGTGAAGCTCAGCCCTTCATCAAT 2100
 DB 66364 TTGTTGATTTACCATCATGATGATTTTACAAAGTGAAGCTCAGCCCTTCATCAAT 66423
 QY 2101 GAGATCTCTGAGTATGATGATTTTAAATGAAGATGATTCCTCTACCTAATCAACA 2160
 DB 66424 GAGATCTCTGAGTATGATGATTTTAAATGAAGATGATTCCTCTACCTAATCAACA 66483

RESULT 2
 LOCUS ATAC011436 106688 bp DNA linear PLN 24-JAN-2001
 DEFINITION Arabidopsis thaliana chromosome III BAC F3L24 genomic sequence,
 complete sequence.
 ACCESSION AC011436
 VERSION AC011436.7 GI:12408733
 KEYWORDS HTG.
 SOURCE Arabidopsis thaliana.
 ORGANISM Arabidopsis thaliana.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 106688)
 Lin, X., Kaul, S., Town, C.D., Benito, M.-I., Creasy, T.H., Haas, B.,
 Roman, C.L., White, O., Nierman, W.C. and Fraser, C.M.,
 Arabidopsis thaliana chromosome III BAC F3L24 genomic sequence
 Unpublished
 2 (bases 1 to 106688)
 Lin, X. and Kaul, S.
 Direct Submission
 Submitted (06-OCT-1999) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA, xlin@tigr.org
 3 (bases 1 to 106688)
 Lin, X.
 Direct Submission
 Submitted (24-JAN-2001) The Institute for Genomic Research, 9712

COMMENT

Medical Center Dr., Rockville, MD 20850, USA
 On Jan 24, 2001 this sequence version replaced gi:12280819.
 Address all correspondence to:
 Xiaoying Lin

The Institute for Genomic Research
 9712 Medical Center Dr.
 Rockville, MD 20850, USA
 e-mail: xlin@tigr.org

BAC clone F3L24 is from Arabidopsis chromosome III and is near the
 molecular marker C1C7A12.
 The orientation of the sequence is from SP6 to T7 end of the BAC
 clone.

Genes were identified by a combination of three methods: Gene
 prediction programs including GBAII (available by anonymous ftp
 from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of
 Washington), Genscan (Chris Burge,
<http://www.cbs.dtu.dk/netgene/obsnetgene.html>), and NetPlantgene
<http://www.stanford.edu/~chris/GENSCANW.html>), searches of the
 complete sequence against a peptide database and the Arabidopsis
 EST database at TIGR (<http://www.tigr.org/tdb/at.html>).
 Annotated genes are named to indicate the level of evidence for
 their annotation. Genes with similarity to other proteins are named
 after the database hits. Genes without significant peptide
 similarity but with EST similarity are named as 'unknown' proteins.
 Genes without protein or EST similarity, that are predicted by more
 than two gene prediction programs over most of their length are
 annotated as 'hypothetical' proteins. Genes encoding tRNAs are
 predicted by tRNAscan-SE (Sean Eddy,
<http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are
 identified by RepeatMasker (Arian Smit,
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Regions of
 genomic sequence that are not annotated as genes but have predicted
 exons by GBAII are annotated as misc features.

FEATURES

source

1..106688

/organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="III"
 /map="C1C7A12"
 /clone="F3L24"
 1..21238

misc_feature

/note="Overlap with BAC clone MZB10
 (AC009326:64324..85561)."
 complement(18..122)

misc_feature

/note="exon predicted by xgrail, quality
 marginal_shadowexon"
 <248..>1916

gene

/gene="F3L24.1"
 /note="MZB10.18; nearly identical to hypothetical protein
 GB:AD56331 [Arabidopsis thaliana]"
 join(<248..469,653..769,852..946,1033..1126,1213..1344,
 1419..1523,1612..1662,1743..>1916)

mRNA

/gene="F3L24.1"
 /note="F3L24.1"
 /codon_start=1

CDS

/protein_id="AAFI4017.1"
 /db_xref="GI:6478912"
 /translation="MALSMEEGFSIGSCFAPNPVYLISAPKRNFTLRKRRRL
 RVSASVKEAEASLETRRRIYLESGLDEKYSMTGLDKELOLAFKSSKIRL
 RSNALINEMTQVDFAGFPEVDTPFCANFTSNVNTVVDLPIHOLITPTGYO
 DKYKNKIMSIYHKYAEFPWGGKLGESIKFSPIDVMTFFSSSKRKRALFSAFLFY
 YOAMLEMTIOVREMPSHVANCERQHKILTRAKQNDDEHGLKRLVGAARKELLR
 DFLFNGVDELGRTFTFDYFPEYOTEDGVSDKRSIIIGKSYETPMDLTQFIC"
 complement(1934..2024)

misc_feature

/note="exon predicted by xgrail, quality
 marginal_shadowexon"
 complement(<2362..>3495)

gene

/gene="F3L24.2"
 /note="MZB10.19; identical to hypothetical protein

GB:AD56336 [Arabidopsis thaliana]: contains Pfam profile:
RNA recognition motif. (aka RRM, RBD, or RNP domain)"
complement(join(<2362..2462,2534..2703,3004..3276,
3353..3384,3475..>3495))
/gene="F3124.2"
complement(join(2362..2462,2534..2703,3004..3276,
3353..3384,3475..3495))
/gene="F3124.2"
/codon_start=1
/product="putative RNA-binding protein"
/protein_id="AAFL4040.1"
/db_xref="GI:6478935"
/translation="MDEIANKGLEVKRSRKIRSDIERISVTGYDIEPREVESELK
KIFSPGEITDVHPIETRNSSLSMHAFIYFEGGTADKALQINGSDMGQVYDAADP
PRKEDNFSIYVLGEDAIDKAPKLSGTDMGSGKRAYRFAYHGRNIFHPRRQGR
ICGRPDKPKMAAIKIKTKPKAKKIKKPKMTLE"
4139..4182
/note="exon predicted by xgrall, quality
marginal shadowexon"
complement(4387..4429)
/note="exon predicted by xgrall, quality good_shadowexon"
complement(4353..4736)
/note="exon predicted by xgrall, quality good_shadowexon"
5389..5442
/note="exon predicted by xgrall, quality good"
5748..9002
/gene="F3124.3"
/note="WZB10.20; similar to MORA transposase of maize
Mutator transposon"
/pseudo
9051..9107
/note="exon predicted by xgrall, quality good"
complement(9181..9241)
/note="exon predicted by xgrall, quality
marginal_shadowexon"
9601..9827
/note="exon predicted by xgrall, quality excellent"
complement(9619..9780)
/note="exon predicted by xgrall, quality good_shadowexon"
complement(<10632..>14351)
/gene="F3124.4"
/note="WZB10.21; identical to hypothetical protein
GB:AD56332 [Arabidopsis thaliana]"
complement(join(<10632..10789,10875..11340,11454..11538,
11622..12016,12222..12341,12417..12578,12644..13375,
13467..14351))
/gene="F3124.4"
complement(join(10632..10789,10875..11340,11454..11538,
11622..12016,12222..12341,12417..12578,12644..13375,
13467..14351))
/gene="F3124.4"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAFL4018.1"
/db_xref="GI:6478913"
/translation="MGVYELPKRFLRGLSEDEGRKRIINFTLMRIETIKALDONHL
KOLBSOGFELMGCTHOFSYMLHYILSRQVLDKKEMMVVRAGKIRIRAIIDFAL
VTGLNCSKSGFNKDKARTGRKMAVONFTRTKALFGKERPTSQMVLGKSMGKTK
DPLTRFLSLLDVGLICPTCSRTIIPDHEVEMKDESLKTPWGTSLATYVLA
KORTAKELADPTAIOGFPAHVAVACPAIITYADSVNIIDPTIPVKLYOYVL
GLSINIVHDAITIDSGMDRVTSILNELLAKEIDLSFSEDEEVSLSLIMEE
HPFENHWSGVDANDVEHPSEDEPMADDCNOAKGDDMDMDVDSGAABERRECA
AGGVGIDIGDANRDPYTDKGVSAFADGELQGVDPAGMENNDSAAHGRENAY
IKEVADERKVLKTIQSSNDELKSYFAELVYOKKELVALITRSVRVAGLNRQAG
IVGSCSKVADASHKQVKNLVLVETSSIALDQIRADPGKHKVSOQVEENIGLSEG
SDNPAAAGINVGVSPEKSNMTNREDEDSIQKGKSKSCSDMSNRPLKSEG
DMEHVASIKYVSPTEKLEDDVDIVENPNSLITQEIINRRLPPRGKKTFFSVSY
LRSNRIRDRPIQSSNIENMTSIPVOPQTSRGAGKSHVSSAKVKEOPTNOSPFI
GPRPFVSPSSVYKRLKQMERGVSYLAAGVYLENEFINTEGQTLDESQDILL
VRFIOLRDKRKLPLRDFLPSFSLSELRKQTNKTVGADGASAFYFTSLARVPFMRK
WFAEIDVLYCPQVROHMGVGLIDLKMGVTVLDCNDVLDKQDVEKVEYVHIITMLP
YLIRKGVNGYKMLTORLEPMTVSRPKIQCSSTSLGVIGIASVILLLEHVSNAMEYCGK
LDGKMRDAAKOYALIEFNALSPTERGA"

	misc_feature	14460..14526
	/note="exon predicted by xgrall, quality excellent_shadowexon"	
	15070..15229	
	/note="exon predicted by xgrall, quality marginal"	
	15834..>17202	
	/gene="F3124.5"	
	/note="WZB10.22; identical to hypothetical protein GB:AD56333 [Arabidopsis thaliana]"	
	join(15834..16133,16228..16393,16566..>17202)	
	/gene="F3124.5"	
	join(15860..16133,16228..16393,16566..17202)	
	/gene="F3124.5"	
	/note="unknown protein"	
	/codon_start=1	
	/protein_id="AAFL4019.1"	
	/db_xref="GI:6478914"	
	/translation="MOTLHOSQLONPFAEANNOSSEDPAPKQVQAQMERLNOARVI ADIRGADRILEAFVASQPRHTDPLQFLREDASMRQHODRLIGKLEESGLV ESLRSRSMGLHMDPLVPCDAVAVAMKROLQAGASAVDRIRLAKFTDQRR FPPIIDGKMKPSSSKRHASHLLENGEEDYVDTLPDIOSRLKLVPSKYVYSGR LNLWLRANSPLSGSDPTEASKPLFQSSSKLRSGLOTEVDKTAIVELSPFLRAI VSLPAGSVDPDAVAFSPDEVTSLALAKYTDHVKLFKPSCFSPFLOGGSYLHA RGFSYHYVKHIVYIKMLIYIV" 17250..17318 /note="exon predicted by xgrall, quality excellent" 17412..17553 /note="exon predicted by xgrall, quality excellent"	
	misc_feature	
	misc_feature	
	Query Match	100.0%; Score 2160; DB 8; Length 106688;
	Best Local Similarity	100.0%; Pred. No. 0;
	Matches 2160; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1	GAATCCCCACGTCACGTGACGTGCAATTCACGTCGAGTGGGCCCTTAAGTTGG 60
DB	1	GAATCCCCACGTCACGTGACGTGCAATTCACGTCGAGTGGGCCCTTAAGTTGG 60
QY	61	ACCATGACTCGGAGCGAGTGTGAATTCATGTCGTCGCAATTCGTTGCTCAGCA 120
DB	61	ACCATGACTCGGAGCGAGTGTGAATTCATGTCGTCGCAATTCGTTGCTCAGCA 120
QY	121	AACTGTGAATAATTTATCTCTTTATAGATAAAGATCTGCTTTTTCAGTTTCAGTA 180
DB	121	AACTGTGAATAATTTATCTCTTTATAGATAAAGATCTGCTTTTTCAGTTTCAGTA 180
QY	181	TGAAGAAGAAATTTGAAGAGAGTGTCCGAGAAAGACCTTGTGCTTCAAGTTGTAGTCT 240
DB	181	TGAAGAAGAAATTTGAAGAGAGTGTCCGAGAAAGACCTTGTGCTTCAAGTTGTAGTCT 240
QY	241	TGTTGTAAAGGCTTATCAATGAGTGGTTGGTTTCATATGGGTCATGCTTCAAGCAC 300
DB	241	TGTTGTAAAGGCTTATCAATGAGTGGTTGGTTTCATATGGGTCATGCTTCAAGCAC 300
QY	301	AAACCCACCTGTTCTAATCTCTGCAAGCCCTAATAGATCAATTCACGTTGAGAAGAG 360
DB	301	AAACCCACCTGTTCTAATCTCTGCAAGCCCTAATAGATCAATTCACGTTGAGAAGAG 360
QY	361	AAAGAAAAGATTTACTTACAGTCTGCTGTGTCGTATGAAGAAATTCGACAGTCTGC 420
DB	361	AAAGAAAAGATTTACTTACAGTCTGCTGTGTCGTATGAAGAAATTCGACAGTCTGC 420
QY	421	TTTAAAGAAACCAAGAAAAGAGATGCTTGAACCTTCAATCCACAGGTATAGCAAT 480
DB	421	TTTAAAGAAACCAAGAAAAGAGATGCTTGAACCTTCAATCCACAGGTATAGCAAT 480
QY	481	TACATTTGCTAGTGTAGTGGAGGATTAATTTCTCATTTGTTCTGCTGTAATTTTG 540
DB	481	TACATTTGCTAGTGTAGTGGAGGATTAATTTCTCATTTGTTCTGCTGTAATTTTG 540
QY	541	GGTAAATGATTTGAGTTGTCTATTAGAACCAACAAATTAAGTTTACTGTTATAGACTGC 600
DB	541	GGTAAATGATTTGAGTTGTCTATTAGAACCAACAAATTAAGTTTACTGTTATAGACTGC 600

QY 601 TTATATAGTAAAGTTCAGATTTTGTTCATACAGAACTGTTTCAGAGAAAGTA 660
 |||||
 Db 601 TTATATAGTAAAGTTCAGATTTTGTTCATACAGAACTGTTTCAGAGAAAGTA 660
 QY 661 TAGTACATGACAGAGCTAGATGTAAGACCACTTCAATGCTTGTAAATCTTC 720
 |||||
 Db 661 TAGTACATGACAGAGCTAGATGTAAGACCACTTCAATGCTTGTAAATCTTC 720
 QY 721 AAAGATTACACTTGTAGAGATGAGCAATGAGATGAGCAATGAGATGAGTCA 780
 |||||
 Db 721 AAAGATTACACTTGTAGAGATGAGCAATGAGATGAGCAATGAGATGAGTCA 780
 QY 781 GCAGTCAAACTGATGCTTGTAGTCCATTTCTTCAATGATGATGATGATGTA 840
 |||||
 Db 781 GCAGTCAAACTGATGCTTGTAGTCCATTTCTTCAATGATGATGATGATGTA 840
 QY 841 TCTGCTTACAGCTTGTGACTTTGGGTTTCAATGAGCCTGATGATGATGATGTA 900
 |||||
 Db 841 TCTGCTTACAGCTTGTGACTTTGGGTTTCAATGAGCCTGATGATGATGATGTA 900
 QY 901 TCTGCTTACAGCTTGTGACTTTGGGTTTCAATGAGCCTGATGATGATGATGTA 960
 |||||
 Db 901 TCTGCTTACAGCTTGTGACTTTGGGTTTCAATGAGCCTGATGATGATGATGTA 960
 QY 961 AGTTATGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGTA 1020
 |||||
 Db 961 AGTTATGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGTA 1020
 QY 1021 GTTCTCTTGTAGGAGCTTAATGCTTGTGATGATGATGATGATGATGATGATGTA 1080
 |||||
 Db 1021 GTTCTCTTGTAGGAGCTTAATGCTTGTGATGATGATGATGATGATGATGATGTA 1080
 QY 1081 CAAGTATTAATCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGTA 1140
 |||||
 Db 1081 CAAGTATTAATCAAGATGATGATGATGATGATGATGATGATGATGATGATGTA 1140
 QY 1141 ACACCAAAATTAATCAAGATGATGATGATGATGATGATGATGATGATGATGATGTA 1200
 |||||
 Db 1141 ACACCAAAATTAATCAAGATGATGATGATGATGATGATGATGATGATGATGATGTA 1200
 QY 1201 ATTATTTGAGAGCTTCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGTA 1260
 |||||
 Db 1201 ATTATTTGAGAGCTTCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGTA 1260
 QY 1261 TCGCCTTTGAGAGCTTCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGTA 1320
 |||||
 Db 1261 TCGCCTTTGAGAGCTTCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGTA 1320
 QY 1321 TCTGCTTGTAGAGCTTCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGTA 1380
 |||||
 Db 1321 TCTGCTTGTAGAGCTTCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGTA 1380
 QY 1381 GAAACTTGTAGAGCTTCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGTA 1440
 |||||
 Db 1381 GAAACTTGTAGAGCTTCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGTA 1440
 QY 1441 AAGTGAAGGAGAGATGAGCAATGATGAGAGCAATGATGAGAGCAATGATGAGAGTA 1500
 |||||
 Db 1441 AAGTGAAGGAGAGATGAGCAATGATGAGAGCAATGATGAGAGCAATGATGAGAGTA 1500
 QY 1501 AACTGACATGAGAGCAATGATGAGAGCAATGATGAGAGCAATGATGAGAGTA 1560
 |||||
 Db 1501 AACTGACATGAGAGCAATGATGAGAGCAATGATGAGAGCAATGATGAGAGTA 1560
 QY 1561 ACAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGTA 1620
 |||||
 Db 1561 ACAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGTA 1620
 QY 1621 CATGCTTCTTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGTA 1680
 |||||
 Db 1621 CATGCTTCTTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGTA 1680
 QY 1681 CATGCTTCTTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGTA 1740
 |||||

Db 1681 CATGCTTCTTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGTA 1740
 |||||
 QY 1741 AGAGCTGCTAAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGTA 1800
 |||||
 Db 1741 AGAGCTGCTAAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGTA 1800
 QY 1801 TTGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGTA 1860
 |||||
 Db 1801 TTGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGTA 1860
 QY 1861 TTGGGAAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGTA 1920
 |||||
 Db 1861 TTGGGAAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGTA 1920
 QY 1921 GATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGTA 1980
 |||||
 Db 1921 GATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGTA 1980
 QY 1981 AGTGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGTA 2040
 |||||
 Db 1981 AGTGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGTA 2040
 QY 2041 TTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGTA 2100
 |||||
 Db 2041 TTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGTA 2100
 QY 2101 GAGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGTA 2160
 |||||
 Db 2101 GAGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGTA 2160

RESULT 3
 AY099706
 LOCUS
 DEFINITION
 Arabidopsis thaliana unknown protein (At3g09150) mRNA, complete cds.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Arabidopsis thaliana.
 Arabidopsis thaliana.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE
 1 (bases 1 to 1358)
 Nguyen,M., Karlin-Neumann,G., Southwick,A., Lam,B., Miranda,M., Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H., Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamiya,A., Kawal,J., Kim,C., Liu,J., Liu,S.X., Narusaka,M., Pham,P.K., Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K., Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.M.
 Direct Submission
 Submitted (24-APR-2002) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
 COMMENT
 e-mail for correspondence: arab@sequence.stanford.edu
 RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : RIKEN Arabidopsis Full-Length cDNA) : Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawal,J., Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Nguyen,M., Southwick,A., Karlin-Neumann,G., Lam,B., Miranda,M., Palm,C.J., Bowser,L., Jones,T., Banh,J., Chen,H., Cheuk,R., Chung,M.K., Kim,C., Lin,J., Liu,S.X., Pham,P.K., Sakano,H., Shinn,P., Yamada,K., Ecker,J., Theologis,A. and Davis,R.M.
 Nguyen,M. (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Davis,R.M.

(SSP/Stanford) contributed equally to this work as PIs.

FEATURES

Location/Qualifiers

1. 1358

/organism="Arabidopsis thaliana"

/db_xref="taxon:3702"

/chromosome="3"

/clone="RAF109-83-B22 (R25336)"

/note="This clone is in pBluescriptpt

ecotype: Columbia"

1. 1358

/gene="At3g09150"

/note="F3L24.1"

139. 1128

/gene="At3g09150"

/codon_start=1

/product="unknown protein"

/protein_id="AAW2057.1"

/db_xref="GI:20466480"

BASE COUNT 426 a 231 c 308 g 393 t

ORIGIN

Query Match

28.1% Score 608; DB 8; Length 1358;

Best Local Similarity 66.7% P-adj. 1e-121; Mismatches 1357; Conservative 0; Indels 679; Gaps 7;

Matches 1357; Conservative 0; Mismatches 0; Indels 679; Gaps 7;

111 GTCCTCACTGAACAGTGTGAAATTTATCTCTTTATAGATTAAGATCTGCTTTTCA 170

Db 2 GTCTCACTGAACAGTGTGAAATTTATCTCTTTATAGATTAAGATCTGCTTTTCA 61

Qy 171 GTTTTCAGTATGAAGAATTTGAAGAGAGTCCGAGAGAGAGACCTTTGTTTCA 230

Db 62 GTTTTCAGTATGAAGAATTTGAAGAGAGTCCGAGAGAGAGACCTTTGTTTCA 121

Qy 231 TTGAGAGCTGCTGTGTAATGCTTTATCAATGAGATTGGGTTTCAATTGGGTCAT 290

Db 122 TTGAGAGCTGCTGTGTAATGCTTTATCAATGAGATTGGGTTTCAATTGGGTCAT 181

Qy 291 TCAAGGACCAAAACCCACCTGTTCAATCTCGAAGCCCTAATAGATCAATTTCA 350

Db 182 TCAAGGACCAAAACCCACCTGTTCAATCTCGAAGCCCTAATAGATCAATTTCA 241

Qy 351 TGAGAGAGAGAGAGAGAGATTTCTACTTAGAGTCTGCTGTGCTGTAAGAGATTC 410

Db 242 TGAGAGAGAGAGAGAGAGATTTCTACTTAGAGTCTGCTGTGCTGTAAGAGATTC 301

Qy 411 CAGAGTCTGCTTTGAGAGAGAGAGAGAGATTCCTTGAACCTTCACATCTCCAG 470

Db 302 CAGAGTCTGCTTTGAGAGAGAGAGAGAGATTCCTTGAACCTTCACATCTCCAG 357

Qy 471 TATATGCAATTAACATTTCTAGTGTAGTGGAGAGATTAATTTCTCTTTCTGCT 530

Db 358 ----- 357

Qy 531 GTGATTTTGGGTAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 590

Db 358 ----- 357

Qy 591 TATAGACTGCTTATATAGTAAAGTTCAATTTGTTTCTTCTAATCAGAAACCTGTT 650

Db 358 -----C 358

Qy 651 AGAGAAAGTATAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 710

Db 359 AGAGAAAGTATAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 418

Qy 711 TTAATCTTCAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 770

Db ----- 770

Db 419 TTAATCTTCAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 475

Qy 771 TTAATCTTCAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 830

Db 476 ----- 475

Qy 831 ATTGTTGTATCTGCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 890

Db 476 -----AGTCTTTTACCTTTTGGGGTTTTCATGAGAGAGAGAGAGAGAG 516

Qy 891 ACTCCATATTTCTGCTTACCTTTTCAATCTTCAATCTTCAATCTTCAATCTTCA 950

Db 517 ACTCCATATTTCTGCTTACCTTTTCAATCTTCAATCTTCAATCTTCAATCTTCA 572

Qy 951 GTTATCTTCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1010

Db 573 ----- 572

Qy 1011 TTTACTGATGTTCTTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1070

Db 573 -----GAGCTTTTATCTTTTCAATCTTCAATCTTCAATCTTCAATCTTCA 610

Qy 1071 ATTACCAAGAGAGATTAATTAACAGATTAATTAACAGATTAATTAACAGATTA 1130

Db 611 ATTACCAAGAGAGATTAATTAACAGATTAATTAACAGATTAATTAACAGATTA 664

Qy 1131 CCACAGAGATTAACAGATTAATTAACAGATTAATTAACAGATTAATTAACAGAT 1190

Db 665 ----- 664

Qy 1191 TGATCTGAGATTAATTTGACAGATTTCCATTTGAGAGAGAGAGAGAGAGAGAG 1250

Db 665 -----AGCTTTTCCATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 704

Qy 1251 AAAGTTTCTGCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1310

Db 705 AAAGTTTCTGCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 764

Qy 1311 GCTTTTCTGCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1370

Db 765 GCTTTTCTGCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 795

Qy 1371 GCTTTTCTGCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1430

Db 796 -----CAGCATGCTTTGAG 810

Qy 1431 ATGACATCAAG 1490

Db 811 ATGACATCAAG 870

Qy 1491 CAACAG 1550

Db 871 CAACAG 901

Qy 1551 GCATGTTGAACAG 1610

Db 902 -----A 902

Qy 1611 GCATGTTGAACAG 1670

Db 903 GCATGTTGAACAG 952

Qy 1671 GATTTGATCCATTTGATGATGATGATGATGATGATGATGATGATGATGATG 1730

Db 953 ----- 952

Qy 1731 GAACAAATCAAG 1790

Db 953 -----AG 1002

Qy 1791 AAACATTCATTTGATTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1850

Db 1003 AAACATTCATTTGATTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1062

QY	1851	CGAAGTATCATTTGGAGACGATATGAACACGTCGCCATGGATTTAACAAGCAATTTATC	1910
Db	1063	CGAAGTATCATTTGGAGAGCTCATATGAACATCTGCTCATGGATTTAACAAGCAATTTATC	1122
QY	1911	GGCTAACAAATGATTTATGTGTGAAACAAGTCAGATTTTCAGAGTCATCAACACAAGACGCTG	1970
Db	1123	GGCTAACAAATGATTTATGTGTGAAACAAGTCAGATTTTCAGAGTCATCAACACAAGACGCTG	1182
QY	1971	AACCTAAGGAAGTGAAGAAATAGAAGACGACATGAGAGAGCTCTCAGGCTATCTGCAT	2030
Db	1183	AACCTAAGGAAGTGAAGAAATAGAAGACGACATGAGAGAGCTCTCAGGCTATCTGCAT	1242
QY	2031	TTCAAGATGATTTGTTGATTTACCATGCATTTGATTTTACAACTGATAGCTTCAGCCCT	2090
Db	1243	TTCAAGATGATTTGTTGATTTACCATGCATTTGATTTTACAACTGATAGCTTCAGCCCT	1302
QY	2091	TCATCAAAATGAGAAATCTCGAGTATGATATGATTTTAATGAAATGATTTGCTT	2146
Db	1303	TCATCAAAATGAGAAATCTCGAGTATGATATGATTTTAATGAAATGATTTGCTT	1358
RESULT 4			
AB045112		1337 bp	linear
LOCUS	AB045112		PLN 03-MAR-2001
DEFINITION	Arabidopsis thaliana mRNA for phytochromobilin synthase HY2		
ACCESSION	AB045112		
VERSION	AB045112.1	GI:13359272	
KEYWORDS			
SOURCE	Arabidopsis thaliana cDNA to mRNA.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Eudicotyledons: core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
AUTHORS	1 (sites) Kohchi,T., Mukougawa,K., Frankenberger,N., Masuda,M., Yokota,A. and Lagarias,J.C.		
TITLE	The arabidopsis hy2 gene encodes phytochromobilin synthase, a ferredoxin-dependent biliverdin reductase		
JOURNAL	Plant Cell 13 (2), 425-436 (2001)		
MEDLINE	21124703		
REFERENCE	2 (bases 1 to 1337)		
AUTHORS	Kohchi,T.		
TITLE	Direct Submission		
JOURNAL	Submitted (21-JUN-2000) Takayuki Kohchi, Nara Institute of Science and Technology, Graduate School of Biological Sciences; 8916-5 Takayama, Ikoma, Nara 630-0101, Japan		
	(E-mail: kouchi@bs.ais-t-nara.ac.jp, Tel:81-743-72-5561, Fax:81-743-72-5569)		
FEATURES			
source	1..1337	Location/Qualifiers	
gene	/organism="Arabidopsis thaliana"		
5'UTR	/db_xref="taxon:3702"		
1..1337			
1..95	/gene="HY2"		
96..1085	/gene="HY2"		
CDS	/codon_start=1		
	/product="phytochromobilin synthase HY2 protein"		
	/protein_id="BAB31374.1"		
	/db_xref="GI:13359273"		
	/translation="MALSMEFSGSCSKAPNPVYLISAPNKINFTIRKKRFLRVSAYSEKFAESALAEETRRKRVLEPSHLOERYSSMTGIDGTELOMLAFKSKRLRMALENEMTVQDFAGMEPEYDPIFCANFETSTNNVIVLDLNPQLQDIDYDLYKYMISIVKRYAETEPWGGKLTGESIKFSPILMMWTRFSSSEKKRAALPSALEYQALMEMTLOVREEMPSHVANRCEAQHRYLTWRAQKDGSHLLKRLVGEARAKELLDLFNGVDELGRTTFIDYRPEYQTEGTVSDKRSIIIGKSYETRRPMDLTNGQFTG"		
	1086..1316		
	/gene="HY2"		
	1316		

BASE COUNT	434 a	226 c	301 g	376 t	
ORIGIN	/gene="HY2" /note="21 a nucleotides"				
Query Match	26.3%; Score 567.2; DB 8; Length 1337;				
Best Local Similarity	65.8%; Pred. No. 7,5e-113;				
Matches 1321; Conservative	0; Mismatches 8; Indels 679; Gaps 7;				
QY	153	AGAACTCTGCTTTTTCACGTTTTCACGATTAAGAGAAATGAGAGAGTCCGAGAGG	212		
DB	1	AGAACTCTGCTTTTTCACGTTTTCACGATTAAGAGAAATGAGAGAGTCCGAGAGG	60		
QY	213	GAGACCTTTGGTTTCAGTTTGTGAGCTTTGTATGCTTATCAAGAGAGTTGGGT	272		
DB	61	GAGACCTTTGGTTTCAGTTTGTGAGCTTTGTATGCTTATCAAGAGAGTTGGGT	120		
QY	273	TTTCAATTGGTCAATGCTTCAAGGCACCAACCACCTGTTCAATCTCGCAAGCCCTA	332		
DB	121	TTTCAATTGGTCAATGCTTCAAGGCACCAACCACCTGTTCAATCTCGCAAGCCCTA	180		
QY	333	ATAAGATCAATTTACGCTTGAGAGGAGAAAGAAAGATTCTTACTAGAGTCTGCTG	392		
DB	181	ATAAGATCAATTTACGCTTGAGAGGAGAAAGAAAGATTCTTACTAGAGTCTGCTG	240		
QY	393	TGTGCTATTAAGAAATTTGGCAGAGCTCTCTTTAGAGAAACCAGAGAAAGATGCTTGG	452		
DB	241	TGTGCTATTAAGAAATTTGGCAGAGCTCTCTTTAGAGAAACCAGAGAAAGATGCTTGG	300		
QY	453	AACCTTCACATCTCCAGGATATATCATTAATTCATTTGCTAGTGTAGTGGAGAGATTATAT	512		
DB	301	AACCTTCACATCTC-----	314		
QY	513	TTTCATTTGTTCTTCTGCTGCTGATTAATTTGGGTAATTTGATTTGATTTGATTTAGAACCA	572		
DB	315	-----	314		
QY	573	AACAAATTAACCTTACTGTTATAGACGCTGTTATTAAGTAAAGTTTCAGATTTGTTTTC	632		
DB	315	-----	314		
QY	633	TAAATCAGAAACTGTTTTCAGAAAAAGTATAGATGACGAGCACTAGATGTAAGACG	692		
DB	315	-----CAGAAAAAGTATAGATGACGAGCACTAGATGTAAGACG	357		
QY	693	AACCTCAATGCTGCTTTTAAATCTTCAAGATTTAGACTTTGAGAGATATGSCAATAG	752		
DB	358	AACCTCAATGCTGCTTTTAAATCTTCAAGATTTAGACTTTGAGAGATATGSCAATAG	417		
QY	753	AGATGAGACAAATGAGGTTTAACCTCAGCAGTACAAACGATATGCTTATCCCATATTC	812		
DB	418	AGAAAGAGACAAATG-----	432		
QY	813	CTTACTTTCAATTGATGATGTTGTATCTTCGCTTAAGTACCTTTGACCTTTGCGGGTTTC	872		
DB	433	-----AGGCTTTGACTTTGCGGGTTTC	455		
QY	873	ATGAGACCTGAGTATGATTAATCTCCCAATTTCTGTCTAATCTTTTACATTTACCAAGCTT	932		
DB	456	ATGAGACCTGAGTATGATTAATCTCCCAATTTCTGTCTAATCTTTTACATTTACCAAGCTT	515		
QY	933	AACATAGTTGTATTTGAAGTTATCTTTAGTTATCTGAGTATACAGTCTGATTTGTC	992		
DB	516	AACATAGTTGATTT-----	529		
QY	993	CAAACTGATGTTCAATATTTTACTGTATGTTCTTCTTAGGACCTTAATCTTTGCATTC	1053		
DB	530	-----GACCTTAATCTTTGCATTC	549		
QY	1053	AGTTGACTGACAGAGGATTTACCAAGACAAAGTATTTACCAAGATATGTCATATATTC	1112		
DB	550	AGTTGACTGACAGAGGATTTACCAAGACAAAGTATTTACCAAGATATGTCATATATTC	609		

```

QY 1113 ACAAAATGCTGAGGTGACCAACAAGATACCAAAATTACTCAATTCGAAGTAAACCTAA 1172
    |||||
Db 610 ACAAAATGCTG----- 621
QY 1173 TCGTGAAGGTAAATGACTGATCTTGAGATTTATTTCCAGACTTCCCATGGGAGGAA 1232
    |||||
Db 622 -----AGACTTCCCATGGGAGGAGAA 643
QY 1233 ATTGACTGTTGAATCCATAAGTTTCTCGCCCTTGGTATGAGGACTAGGTTTGTGTC 1292
    |||||
Db 644 ATTGACTGTTGAATCCATAAGTTTCTCGCCCTTGGTATGAGGACTAGGTTTGTGTC 703
QY 1293 TAGCAAGAAAGAAACATTAAGGCTTGTCTCTGCTTTCTAGAGTACTATCAGTATATAC 1352
    |||||
Db 704 TAGCAAGAAAGAAACATTAAGGCTTGTCTCTGCTTTCTAGAGTACTATCAGTATATAC 752
QY 1353 TCAGCGGCCAAAAGCTAAGTTTATTTGAAACTTTGACATGACATATATCATCTCTTC 1412
    |||||
Db 753 ----- 752
QY 1413 CTACAGGCGATGGCTTGAGATGACAATCCAGTGAAGGAGATGGAACCATCTCATGTG 1472
    |||||
Db 753 -----CAGGCGATGGCTTGAGATGACAATCCAGTGAAGGAGATGGAACCATCTCATGTG 809
QY 1473 AGAGCCAAATTTGTGAAGCACAACAACATGACTGACATGGGAGACAAAAGGTGATTTC 1532
    |||||
Db 810 AGAGCCAAATTTGTGAAGCACAACAACATGACTGACATGGGAGACAAA----- 858
QY 1533 TTTTCCTTTTGTATATTGATGATGTTGACAGACATCTATCTGATTTTACATGATGAT 1592
    |||||
Db 859 ----- 858
QY 1593 ATTGATTTGGTGTTCGAGGATCTGAGCATGCTCTTAAAGATTAGAGTGAAGC 1652
    |||||
Db 859 -----AGATCTGAGCATGCTCTTAAAGATTAGAGTGAAGC 901
QY 1653 AAAGCAAGATATTAAGATTTGATCCATTAAGTGTCCCATTTATTAAGTGTGTA 1712
    |||||
Db 902 AAAGCAAA----- 909
QY 1713 AGATGTTGAAATGATTTGACAACAATCAGAGGCTGCTAAAGGATTTCTGTTCAATGG 1772
    |||||
Db 910 -----AGAGGCTGCTAAAGGATTTCTGTTCAATGG 941
QY 1773 GTGGATGATTTAGGACAAAACATCATTTGATTTCTCCAGAGTACCAAGCAGATGAT 1832
    |||||
Db 942 GTGGATGATTTAGGACAAAACATCATTTGATTTCTCCAGAGTACCAAGCAGATGAT 1901
QY 1833 GGAAGCTGTAAGCATAAAGCATATGGAAGTCAATATGAACCTGTCATGGGAT 1892
    |||||
Db 1002 GGAAGCTGTAAGCATAAAGCATATGGAAGTCAATATGAACCTGTCATGGGAT 1961
QY 1893 TTTAAGGAGCAATTTATCGGCTAATCATATATGGAACAGTCAAGATTTCAAGATCA 1952
    |||||
Db 1062 TTTAAGGAGCAATTTATCGGCTAATCATATATGGAACAGTCAAGATTTCAAGATCA 1121
QY 1953 TCAACACAAGAGACGGAATTAAGGAGTGAAGTAAAGAAAGAGCAGATGGAGTGC 2012
    |||||
Db 1122 TCAACACAAGAGACGGAATTAAGGAGTGAAGTAAAGAAAGAGCAGATGGAGTGC 1181
QY 2013 TCTCAGGCTATCTGATTTCAAGATGATTTGAGTATCAATGATGATTTGATTTACA 2072
    |||||
Db 1182 TCTCAGGCTATCTGATTTCAAGATGATTTGAGTATCAATGATGATTTGATTTACA 1241
QY 2073 AGGTGAGCTCTCAGCCCTTATCAAAATGAGATCTCGAGTATGATATGATTTAATGA 2132
    |||||
Db 1242 AGGTGAGCTCTCAGCCCTTATCAAAATGAGATCTCGAGTATGATATGATTTAATGA 1301
QY 2133 AAATGATTTGCTCTACTATATCAACA 2160
    |||||
Db 1302 AAATGATTTGCTCTCAAAAAAAAAAAAA 1329

```

```

RESULT 5
AY088874
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Arabiidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 1308)
Haas,B.J., Volkovskiy,N., Town,C.D., Tzoukhan,M., Alexandrov,N.,
Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.
Full-length messenger RNA sequences greatly improve genome
annotation
JOURNAL
Genome Biol. (2002) In press
REFERENCE
2 (bases 1 to 1308)
Brover,V., Tzoukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
Feldmann,K.
Full-length cDNA from Arabidopsis thaliana
JOURNAL
Unpublished
AUTHORS
3 (bases 1 to 1308)
Brover,V., Tzoukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
Feldmann,K.
Direct Submission
Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road,
Malibu, CA 90265, USA
COMMENT
This clone sequence is one of 5,000 Ceres full-length cDNAs made
available to TIGR and Genbank. The following quality assessment of
this set was done by comparison with known proteins: two percent
of the clones are estimated to be 5'-truncated; less than one percent
are 3'-truncated; approximately two percent represent alternative
splice variants, including unspliced introns and spliced exons; one
percent may contain premature stop codons; five percent may have
frame shifts in a coding region. A sequence is considered to be
5'-truncated if it lacks the translation initiation start (ATG). A
sequence is considered to be 3'-truncated if it lacks the
C-terminal end of the encoded protein. Please note that these cDNA
sequences are derived from the WS or Ler ecotypes and therefore
may contain polymorphisms when compared to sequences from Col-0.
Geneset carried out the library production and sequencing of the
full-length clones. Ceres, Inc. carried out the clustering of the
5' sequences, selection of clones, and sequence assembly.
FEATURES
Source
Location/Qualifiers
1..1308
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="98026"
82..1053
/codon_start=1
/product="unknown"
/protein_id="AA067180.1"
/db_xref="GI:21618130"
/translational="MERGFSIGSCFKAPNPVLLSASPNKINFLRRNRKFLRVSVA
VSKPEFASLEETFRKRIIVLEPSLQVSNMGLDGLKELQMLAKSKIRILRSMALR
NETNOVDFRAGFERPEYDPTFCANFTSTVNVIVVDLNLPHLDTQDPTQDYK
IMSTYHRYAEFFPWGKLVGTSIKFFSPVLMWTFSSKRLKLVGAKAKELRDLFNG
MTLQVREMERSESHVRANCAEOHKYLTFRAPQDPGHLKRLKLVGAKAKELRDLFNG
VDELGRTKFTFDYPEYOTEDGYVSDKSIISIKSYETRPMDLTGQFIG"
BASE COUNT
414 a 226 c 296 g 372 t
ORIGIN
Query Match 24.3%; Score 525.2; DB 8; Length 1308;
Best Local Similarity 65.2%; Pred. No. 9.9e-104;
Matches 1293; Conservative 0; Mismatches 3; Indels 686; Gaps 8;
QY 180 ATGAAGAAGATTGAAGAGAGTCCGAGGAAGAGACCTTGG-TTTCAGTTGTAGT 238
    |||||
Db 1 ATGAAGAAGATTGAAGAGAGTCCGAGGAAGAGACCTTGGTTTTCAGTTGTAGT 60
QY 239 CTGTGTAAATGCTTATCATGAGATTTGGCTTTCAATTGGTCATGCTTCAAGCA 298

```



```

Db      |||||||
QY      61 CTGTGTGATGCTTATCAATGAGATTGGTTTCAATGGGTGCTAGCTTCAAGCA 120
Db      |||||||
QY      299 CCAAGCCACCGTCTCTATCTCTGCAAGCCCTAATTAAGATCAATTTACGTTGAGAG 358
Db      |||||||
QY      121 CCAAGCCACCGTCTCTATCTCTGCAAGCCCTAATTAAGATCAATTTACGTTGAGAG 180
QY      359 AAGAAAGAAAGATTTCTTACTTAGAGTCTCTGCTGCTGCTAATAGAAATTCGACAGTCT 418
Db      |||||||
QY      181 AAGAAAGAAAGATTTCTTACTTAGAGTCTCTGCTGCTGCTAATAGAAATTCGACAGTCT 240
QY      419 GCTTTAGAAAGAAAGCAAGAAAGATTCCTTGAACCTTCACATCTCCAGGTATATGCA 478
Db      |||||||
QY      241 GCTTTAGAAAGAAAGCAAGAAAGATTCCTTGAACCTTCACATCTCCAGGTATATGCA 289
QY      479 ATTACATTTCTGTTAGTGTGAGAGATTATATTTCTCATTTCTTCTGCTGCTGATTT 538
Db      |||||||
QY      290 ----- 289
QY      539 TGGGTAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 598
Db      |||||||
QY      290 ----- 289
QY      599 GCTTATATAAGTAAAGTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 658
Db      |||||||
QY      290 ----- 291
QY      659 TATAGATGATGACAGAGATAGATGATGATGATGATGATGATGATGATGATGATGATGAT 718
Db      |||||||
QY      292 TATAGATGATGACAGAGATAGATGATGATGATGATGATGATGATGATGATGATGATGAT 351
QY      719 TCAAGATTTAGATCTTTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 778
Db      |||||||
QY      352 TCAAGATTTAGATCTTTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 400
QY      779 CAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 838
Db      |||||||
QY      401 ----- 400
QY      839 TATCTGCTTGTAGTCTTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 898
Db      |||||||
QY      401 ----- 449
QY      899 ATTCTGCTTGTAGTCTTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 958
Db      |||||||
QY      450 ATTCTGCTTGTAGTCTTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 497
QY      959 CTAGTATGCTGAGATTATGAGGTCTGTATTTGATTTGATTTGATTTGATTTGATTTGATTT 1018
Db      |||||||
QY      498 ----- 497
QY      1019 ATGTTCTTCTTTAGGAGCTTATCTTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 1078
Db      |||||||
QY      498 ----- 543
QY      1079 GACAAGTATTAACAAGATATGTCATATATCAACAATATGCTGAGGTGACCAACAAGA 1138
Db      |||||||
QY      544 GACAAGTATTAACAAGATATGTCATATATCAACAATATGCTGAGGTGACCAACAAGA 589
QY      1139 ATACACCAATTTACTCAATTCGAAGTAACTAATGCTGAGGTGATTAATGATGATCTTG 1198
Db      |||||||
QY      590 ----- 589
QY      1139 AGATTTATTTGACAGCTTTCCATGGGAGGAAATTTGATTTGATTTGATTTGATTTGATTTGATTT 1258
Db      |||||||
QY      590 ----- 637
QY      1259 TCTGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1318
Db      |||||||
QY      638 TCTGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 697
QY      1319 TCTGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1378

```

```

Db      698 TCTGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
QY      1379 TGAAGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1438
Db      721 -----
QY      1439 CCAAGTGGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1498
Db      744 CCAAGTGGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 803
QY      1499 GTACCTGACATGCGAGACACAAAGATGATTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 1558
Db      804 GTACCTGACATGCGAGACACAA----- 826
QY      1559 GACGACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1618
Db      827 ----- 835
QY      1619 GACATGCTTCTTTTAAAGATTTAGTGTAGTGAAGCAAGCAAGGATTTAAAGATTTGAT 1678
Db      836 GACATGCTTCTTTTAAAGATTTAGTGTAGTGAAGCAAGGCA----- 877
QY      1679 CCATTTAGTGTGCTTCTTTTAAAGATTTAGTGTGAAGATGTTGAAGATTTGAACAA 1738
Db      878 ----- 877
QY      1739 TCAGAGCTGCTTAAGGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1798
Db      878 --AGAGCTGCTTAAGGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 935
QY      1799 CATTTATTTCTTTCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1858
Db      936 CATTTATTTCTTTCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 995
QY      1859 CATTTGGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1918
Db      996 CATTTGGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1055
QY      1919 ATGATATATGTAAGCAAGTCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1978
Db      1056 ATGATATATGTAAGCAAGTCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1115
QY      1979 GAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2038
Db      1116 GAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1175
QY      2039 GATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2098
Db      1176 GATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1235
QY      2099 ATGAGATCTGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2158
Db      1236 ATGAGATCTGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1295
QY      2159 CA 2160
Db      1296 CA 1297

RESULT 6
AY128900 1092 bp mRNA linear PLN 07-AUG-2002
LOCUS Arabidopsis thaliana unknown protein (Atg90150) mRNA, complete
DEFINITION cds.
ACCESSION AY128900
VERSION AY128900.1 GI:22136443
KEYWORDS FLI CDNA.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1092)

```


misc_feature complement(3872..3962)
/note="exon predicted by xgrall, quality
marginal_shadowexon"
repeat_region complement(3922..4055)
/rpt_family="(GAAAA)n"
gene <4496..>5463
/gene="MZB10.2"
/note="similar to ylippee putative zinc binding protein
GB:RAD47882 [Homo sapiens]"
mRNA join(<4496..4591,5070..5113,5199..5307,5386..>5463)
/gene="MZB10.2"
CDS join(4496..4591,5070..5113,5199..5307,5386..5463)
/gene="MZB10.2"
/codon_start=1
/product="ylippee-like protein"
/protein_id="AAD56315.1"
/db_xref="GI:5923664"
/translation="MGRFLVLDLGLVYSCRCQTHPAVNDIISKSFHCKHGAAYLF
DNVNVTVGEKEHRVMTGMHTVADIFCVSGSLVGMKYEIAYDKSQYKKGKFI
FSSFNV"
misc_feature complement(5490..5586)
/note="exon predicted by xgrall, quality
marginal_shadowexon"
misc_feature 5334..5615
/note="exon predicted by xgrall, quality excellent"
misc_feature complement(5641..5725)
/note="exon predicted by xgrall, quality
excellent_shadowexon"
misc_feature complement(6084..6095)
/note="exon predicted by xgrall, quality
marginal_shadowexon"
misc_feature 6393..6476
/note="exon predicted by xgrall, quality excellent"
misc_feature complement(6418..6474)
/note="exon predicted by xgrall, quality
excellent_shadowexon"
repeat_region complement(6632..6705)
/rpt_family="(GA)n"
gene <6920..>9232
/gene="MZB10.3"
/note="similar to unknown protein GB:AA895292 [Arabidopsis
thaliana]"
mRNA join(<6920..7010,7097..7253,7541..7646,7762..8769,
8969..>9232)
/gene="MZB10.3"
CDS join(6920..7010,7097..7253,7541..7646,7762..8769,
8969..9232)
/gene="MZB10.3"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAD56316.1"
/db_xref="GI:5923665"
/translation="MLTRHDEELSLLEMRREKREHRAADSLTNGSDVSNATLTAA
AAALSGVSEETASQRYPLRRTAAENLYSNEKSDVDWLTTPGCTPQFEKSHRSVA
NODAPNSRPTVLSRLGNCREDIVSGNNKPNQSSSSVAGLRPSSGSSRSRPA
TPTRSTPTPTSTRPVYTRASNRSSTPTSRATLTAAATTTAAAPTTPSSGSA
SATPTRSNPPRSSASKKPVSRPATPTPRRSTPTGSPVSSVARSKAPRSPPTVNS
KAPSRGTPSPPTLSSRRPKPMPGCSLEAPPLRTTLARPYASASGRGVASAR
SRSGSTIEGGPTSGGSGNARQSCSPSRKAPLIGNTIGSLTGVGRKAKNGSGCD
NLSPVAMGNKVERVNNRKLGPRLTENGGSGKSSASNSLGGYGRKAKNGSGCD
LRHMDIRGTMGNLPLVTKVPASMYSVRRPQSPVSSSPVATSSITYSSSDPSVDNI
ILICDGNDAENDDLISRSYASPRNOYPKFTS"
misc_feature complement(10821..10843)
/note="exon predicted by xgrall, quality
marginal_shadowexon"
misc_feature 10828..10857
/note="exon predicted by xgrall, quality marginal"
gene <11221..>12992
/gene="MZB10.4"
/note="similar to receptor kinase GB:S70769 [Arabidopsis
thaliana]"
mRNA join(<11221..11263,11488..11600,11690..11900,11982..12219,
12398..12448,12597..>12992)

CDS

Query Match 4.3%; Score 92.6; DB 8; Length 8561;
Best Local Similarity 76.8%; Pred. No. 4.1e-10;
Matches 179; Conservative 0; Mismatches 44; Indels 10; Gaps 5;

gene

/gene="MZB10.4"
join(11221..11263,11488..11600,11690..11900,11982..12219,
12298..12448,12597..12992)
/gene="MZB10.4"
/codon_start=1
/product="putative receptor ser/thr protein kinase"
/protein_id="AAD56317.1"
/db_xref="GI:5923666"
/translation="MCGSDRLGGREAETCTNNVRFVSNLSASADPSPHTRIGG
GGYGVKGVLRDSTQAVAYKLSLSAESQGRFETETNLINIHHPNLVKLIGCIG
NNRLIVETLENNLSAYLSRSRRTVPLDMSKRAICVGSGLAFIHEVEPHVH
RDIKASNLIDDSNFSPIKIGDGLAKLPEDNVHTVSTVAGTVGLAYLALGQUTK
ADVYSFGILVLEVISGNSSTRAAFGEDEMYLVEMWMLREERLIECDPELTKEPD
EVEFEIKVLECTOAAOKRPMKOVEMMLRKLINLNEALTEPGYRVGRGNRNR
GIGLRVGGSSSSQSRDNRKXSNVPGOSTSNSINISFQSIIEIENLRP"
124213..125448
/gene="MZB10.5"
/note="similar to hypothetical protein GB:AAC27164

RESULT 8

AP003735/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AP003735 150053 bp DNA linear PLN 13-MAR-2002

Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
BAC clone: B1147A04.

AP003735

AP003735.4 GI:19386808

Oryza sativa (japonica cultivar-group) (cultivar: Nipponbare) DNA,
clone: B1147A04.

Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehharitoidae; Oryzae; Oryza.

1

Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
clone: B1147A04
Published Only in Database (2001)
2 (bases 1 to 150053)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submission
Submitted (13-JUN-2001) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: tsasaki@nias.affrc.go.jp, URL: http://rtp.dna.affrc.go.jp/
Tel: 81-298-38-7441, Fax: 81-298-38-7468)
On Mar 12, 2002 this sequence version replaced gi:16904696.
GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
(October 1998 version). The genomic sequence was searched against
NCBI Nonredundant Protein database, nr
(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at


```

Db      10246 GTTGTGTTTTCATGTTTGGTTTGTAGATCTTTGTAATAATTTTGGGCTTTGTTT 10305
QY      565 AGGACCAACAATACTTACTGTATAGACTGCTTATATAGTAAGTAAGTTCAGATT 624
Db      10306 TTGTTGTAATATATTTGATTTAGATTGGAATTTAAGTTTCTTTTGTATATTT 10365
QY      625 TGTGTTTCTAATACAACTGTTTCAGAAAGTATAGACATACAGCACTAGATGG 664
Db      10366 ATTTTGTGTTAGTTAAGTTAAGTAGAAGAAATTTGTGAGATTATAGAGAAAGAA 10425
QY      685 TAAGACCAACTCAATGCTGCTTTTAATCTCAAGATTAGACTGCTGAGAGTAT 744
Db      10426 AAGGTGAATTAATAGTAATAATTTTAATGTTTAT - ATTAATAGTTGGTTTAT 10483
QY      745 GGCATAATAGAGATGAGACAATGCAAGTTTACTTCAGCACTACAACTGATTCCTT 804
Db      10484 TGTGTTGTTGCGGGGTTTTTTTGTGTACCTGATATTTAGCTTATGTTGTTTTT 10543
QY      805 CCATTTCTTACTTCAATGATGATGATGTTGTTGATCTGCTTACGCTTACGCTT 864
Db      10544 CGTTTATTTTGTGTTTATTTTATTTATATATGCTTTTATTTTATTTTATTTT 10603
QY      865 CGGGTTTCATGAGACCTGAGTAT 887
Db      10604 TTTGTTTAGGAATTTTAATGT 10626

```

```

RESULT 11
AX083744
LOCUS      AX083744      1141 bp      DNA      linear      PAT 28-FEB-2001
DEFINITION Sequence 22 from Patent WO0111061.
ACCESSION AX083744
VERSION    AX083744.1 GI:13185472
KEYWORDS
SOURCE     synthetic construct.
ORGANISM   synthetic construct
            artificial sequences.
            1 (bases 1 to 1141)
REFERENCE  1 Kunst, L. and Clemens, S.
            Regulation of embryonic transcription in plants
            Patent: WO 0111061-A 22 15-FEB-2001;
            UNIVERSITY OF BRITISH COLUMBIA (CA)
FEATURES
    source
        1..1141
        location/Qualifiers
            /organism="synthetic construct"
            /db_xref="taxon:32630"
    promoter
        1..1141
        /note="consensus sequence of A.t., L.a., and B.n. FAEI
        promoters"
BASE COUNT 123 a      32 c      42 g      112 t      832 others
ORIGIN

```

```

Query Match      2.28; Score 48.4; DB 6; Length 1141;
Best Local Similarity 10.6%; Pred. No. 3.3;
Matches 105; Conservative 369; Mismatches 515; Indels 4; Gaps 3;
QY      415 GTCTGCTTAGAAGAAACGAGAAAGGATGCTTCTGACCTTCACATCTCCAGGATA 474
Db      81 GWTKKMYBCANNTSBRYHARRKMDKTAIBMTNMGWGTGMHRHRRABDVDDH 140
QY      475 TGCATTTACATTTGCTAGTAGTGGAGATTATTTCTCATTTGTTCTTGCTGTA 534
Db      141 YVTAANNAAWTKCMKDRKRTKRWMMKNNNATGMDDDTYKHMMNNNGCTVYWMRYKT 200
QY      535 ATTTTGGTAAATTTGATTTGATGTCATTAGGAACCAACAATACTTACTGTTATA 594
Db      201 DRDMSBRMYGMBWMKMSYDVYTYWVMDCKRKVRBRVTRGRMRMYVYAWBYAHR 260
QY      595 GACTGCTTATATAGTAAGTAAGTTCATTTGTTTCTTATATACGAACATGTTTCAGCA 654
Db      261 RRYNNGTBAAYRRWTNNNNNNNAKMKRARKARYWGNRABVNSTCTWMSKTKYVTS 320

```

```

QY      655 AAGTATAGTACATGACAGACTAGATGTAAGACCAACTCAATGCTGTTTAA 714
Db      321 WANNCPAGANKDHKWKKSALAMGYVNNNNNNNTYTKARRBARNDWYHMKMKHYN 380
QY      715 ATCTTCAAGATTAGACTCTTGAAGATATGCAATAGAAATGACAACTCAGGTTTA 774
Db      381 AAHSRKKWTBYKRTMNVNNNGTTWKRBMWAMYMDKD -WBGTYNNNNNGRTYYGWT 439
QY      775 ACTTCAGAGTACAAACTATGCTTTTACTCCATTTCTCTTCAATGATGATG 834
Db      440 KKKMWTYYKMANCKWAMPDHKTCTHNTTWWKRYTWNCCYKMSNGSHSRBAAY 499
QY      835 TTGTATCTTCCTTACTGCTTGTGACTTTCGCGGTTTCAAGAGCTGATATGATCTC 894
Db      500 YTYMMWMMRRYAHANNNNNDYMWKACTWYKBYCSWMMNYAAWTKSSWNTSRYRYRK 559
QY      895 CCATATCTGTGCTAATTTTTCACATCTTACCAAGCTTAACATAGTTGATTTGAAGT 954
Db      560 TNSWNRSDTRS -MGRANNYARABHYGYKWTYRMBMSHTWBHBRAGAHAHWMBMXBA 618
QY      955 TCTTCTAGTATGCTGAGATTAACAGTCTGATTTGCAAACTGATCAATATTTTA 1014
Db      619 KCHCMKMYKAKKYAGAGSSNNNNNNNNNNNNNNNNATCARODYTAASKWTAANAAMKY 678
QY      1015 CTGTATGTTCTCTTAGGAGCCTTAATCTTGTGATCATGTTGACTGACCAAGGATTA 1074
Db      679 YKBAANNAYYTHANNWGCWNNATDTRFTWKKNNNNNAGTWKNNNNNAKNA -SAKN 736
QY      1075 CCAAGCAAGTATTAACAGATATATGTCATATATCAAAATATGCTGAGTGACC 1134
Db      737 YAAAYAKAKKHHRMANKMARMGHDAAABTIDKRNNGAYRYKTTTNNNTYRGVYWT 796
QY      1135 AAGAATACACCAATTAATCAATGCAAGTAAACCTAATGCTGAGGTGAATGACTGAT 1194
Db      797 AARDGANNNNNNNNNNNNNNNGSDWVYWMAYANTCTNNNNNNNNNAYAWTKRYTT 856
QY      1195 CTGAGATTTATTTGACAGCTTCCATGCGGAGGAGAAATGACTGATCCATCAAG 1254
Db      857 TDDRMBATYNNNNNNRMAVYGAADYAVYMSDTCDAWMMKDATKNNATYNNRGTAWR 916
QY      1255 TTTTTCGCGCTTGTGATGTGACTAGTGTTCCTCTAGCAAAAGAAACATAAGCT 1314
Db      917 TNNNNNNMKRYHYHAAAMNNNNNNNGKCTATHTWVCAKATYTKGCWNNCTTCKRYN 976
QY      1315 TTGTTCTGCGCTTCTTAGAGCTATCATGATATATACAGCGGCAAAAGTAAAGTT 1374
Db      977 NCTWYMTTTRTWTAAATRKTNATGSTRONATGMMNNNNYTWCKRTWATAYRATW 1036
QY      1375 TTATGGAACCTTGTGACGAGAACTATCATCT 1407
Db      1037 MKAMMKYMATGSWNTNSYARMAYKTRAYKGY 1069

```

```

RESULT 12
AL159996
LOCUS      AL159996      188357 bp      DNA      linear      PRI 24-JUL-2000
DEFINITION Human DNA sequence from clone RP11-401F12 on chromosome 9. Contains
ACCESSION AL159996
VERSION    AL159996.7 GI:8648450
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 188357)
            Laid, G.
            Direct Submission
            Submitted (19-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
            requests: clonerequest@sanger.ac.uk
            On Jun 21, 2000 this sequence version replaced gi:8346239.
            During sequence assembly data is completed from overlapping clones.
COMMENT

```

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr9>

RP11-401F12 is from the library RPCI-11.2 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>

VECTOR: pBACe3.6

This sequence is the entire insert of clone RP11-401F12. The true left end of clone RP11-208G24 is at 167549 in this sequence. The true right end of clone RP11-395D3 is at 5200 in this sequence.

FEATURES

SOURCE

1.188357
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-401F12"
/clone.lib="RPCI-11.2"
410..897
/note="match: GSS: Em:A0606829"
1734..2348
/note="match: GSS: Em:A0823967"
1819..2460
/note="L1MC4 repeat: matches 6462..7077 of consensus"
2551..2614
/note="2 copies 32 mer 96% conserved"
2690..2698
/note="TN10 excised: This sequence represents the duplicated flanking sequence of the TN10."
3529..3560
/note="8 copies 4 mer g1gt 96% conserved"
3746..3959
/note="L1PB3 repeat: matches 5916..6147 of consensus"
3969..4006
/note="U2 repeat: matches 1..38 of consensus"
4392..4430
/note="13 copies 3 mer aac 92% conserved"
5217..5472
/note="match: GSS: Em:A0302480"
5260..6024
/note="match: GSS: Em:A0901451"
complement(5743..6091)
/note="match: GSS: Em:A0132828"
6101..6658
/note="match: GSS: Em:A0830958"
7246..8450
/note="Tigger3b repeat: matches 8..1231 of consensus"
8626..8773
/note="L2 repeat: matches 2452..2611 of consensus"
10391..10584
/note="MER3A repeat: matches 1..189 of consensus"
11880..12174
/note="L2 repeat: matches 1986..2272 of consensus"
complement(12504..12890)
/note="match: STS: Em:HS311YCI"
12609..12648
/note="10 copies 4 mer g1gt 90% conserved"

repeat_region 12610..12647
/note="19 copies 2 mer tg 92% conserved"
repeat_region 12615..12650
/note="6 copies 6 mer g1gtgt 91% conserved"
repeat_region 13805..14099
/note="AluSg repeat: matches 2..295 of consensus"
repeat_region 14101..14388
/note="AluSg repeat: matches 16..302 of consensus"
repeat_region 15605..15839
/note="L2 repeat: matches 2357..2612 of consensus"
repeat_region 15947..16013
/note="L2 repeat: matches 2609..2674 of consensus"
repeat_region 16084..16364
/note="AluJo repeat: matches 1..288 of consensus"
complement(16254..16783)
/note="match: GSS: Em:A0764225"
17436..17528
/note="MIR repeat: matches 35..123 of consensus"
18177..18298
/note="FLAM A repeat: matches 1..122 of consensus"
19380..19675
/note="AluX repeat: matches 1..296 of consensus"
complement(19534..19956)
/note="match: GSS: Em:A0042003"
19736..19958
/note="L2 repeat: matches 2477..2745 of consensus"
20126..20237
/note="L2 repeat: matches 2589..2705 of consensus"
20737..20812
/note="MAD1 repeat: matches 5..80 of consensus"
21025..21200
/note="L2 repeat: matches 2184..2355 of consensus"
21443..21646
/note="MIR repeat: matches 40..246 of consensus"
23474..23793
/note="MIR repeat: matches 197..541 of consensus"
complement(25531..26230)
/note="match: GSS: Em:A0787911"
27777..28867
/note="L1MC3 repeat: matches 6667..7734 of consensus"
28802..28880
/note="L1MC2 repeat: matches 6247..6326 of consensus"
28924..29133
/note="MER20 repeat: matches 2..214 of consensus"
29810..29849
/note="20 copies 2 mer ac 80% conserved"
29868..30157
/note="AluY repeat: matches 3..290 of consensus"
30822..31276
/note="MIR repeat: matches 70..532 of consensus"
31665..32020
/note="THE1B repeat: matches 1..364 of consensus"
32181..32674
/note="match: GSS: Em:A082292"
32279..32403
/note="MIR repeat: matches 57..192 of consensus"
complement(32857..33263)
/note="match: GSS: Em:A0139331"
33279..33551
/note="AluSg repeat: matches 1..301 of consensus"
33349..33588
/note="10 copies 4 mer aac 92% conserved"
35353..35388
/note="3 copies 12 mer 94% conserved"
35759..36002
/note="MIR repeat: matches 2..246 of consensus"
36104..36469
/note="match: GSS: Em:A031409"
37497..37545
/note="MIR repeat: matches 81..127 of consensus"
37664..38248
/note="MIR2D repeat: matches 1..553 of consensus"
38259..38373
repeat_region

[illegible]

*	85211	85214:	contlg of	2494	bp	in	length
*	85215	85314:	gap of	100	bp		
*	85315	87669:	contlg of	2355	bp	in	length
*	87670	87766:	gap of	100	bp		
*	87770	88611:	contlg of	1842	bp	in	length
*	88712	89711:	gap of	100	bp		
*	89712	91891:	contlg of	2180	bp	in	length
*	91892	91991:	gap of	100	bp		
*	91992	93864:	contlg of	1873	bp	in	length
*	93865	93964:	gap of	100	bp		
*	93965	96205:	contlg of	2241	bp	in	length
*	96206	96305:	gap of	100	bp		
*	96306	96749:	contlg of	444	bp	in	length
*	96750	96849:	gap of	100	bp		
*	96850	96800:	contlg of	1951	bp	in	length
*	96801	98900:	gap of	100	bp		
*	98901	100409:	contlg of	1509	bp	in	length
*	100410	100509:	gap of	100	bp		
*	100510	102668:	contlg of	2159	bp	in	length
*	102669	102768:	gap of	100	bp		
*	102769	105539:	contlg of	2771	bp	in	length
*	105540	105639:	gap of	100	bp		
*	105640	108013:	contlg of	2374	bp	in	length
*	108014	108113:	gap of	100	bp		
*	108114	110443:	contlg of	2330	bp	in	length
*	110444	110543:	gap of	100	bp		
*	110544	113588:	contlg of	2845	bp	in	length
*	113589	113488:	gap of	100	bp		
*	113489	115702:	contlg of	2214	bp	in	length
*	115703	115802:	gap of	100	bp		
*	115803	119059:	contlg of	3257	bp	in	length
*	119060	119159:	gap of	100	bp		
*	119150	121502:	contlg of	2343	bp	in	length
*	121503	121602:	gap of	100	bp		
*	121603	123908:	contlg of	2306	bp	in	length
*	123909	124008:	gap of	100	bp		
*	124009	126227:	contlg of	2219	bp	in	length
*	126228	126327:	gap of	100	bp		
*	126329	128037:	contlg of	2710	bp	in	length
*	129038	129137:	gap of	100	bp		
*	129138	131820:	contlg of	2683	bp	in	length

Query Match	2.28;	Score 47.8;	DB 2;	Length 204782;
Best Local Similarity	35.08;	Pred. No. 1.9;		
Matches 109;	Conservative 0;	Mismatches 202;	Indels 0;	Gaps 0;

Oy	513	TTCGATGTCTTGCGTGGAAATTTGGGTAATAATTGAGTTGATCGCATAGGAACCA	572
Dd	25293	TTCTACTGTGCTTTTGTGGCCCATCAGACACATGAATGATGGGATATATTTTCCCTT	25353E
Oy	573	AACAATTAAGTTACTGTATAGACTGCCTATATAGAATAAACCTCAGATTTGTTTTTC	632
Dd	25353	ACCAAATAAGCTGAGACTTCTAGACGCTCTACTGTTCAAATCCTTGGATTTGGCTTTTA	25412E
Oy	633	TAATCACGAACTGTTTCAGGAAAAGTATATGTAGCATGCACAGACTAGATGTTAGACC	692
Dd	25413	TACTTCTTGCTTTTTCATTGCTGTATGTTATGTAGGGAAGACCTCTCCCAATCCYG	25472Z
Oy	693	AACCTCAATGCTTGCCTTTAATCTTCAAGAATTAGACTCTGAGAGATATGGCATAAG	752
Dd	25473	AGTTTGANNN	25532Z
Oy	753	-AGATGAGACAATGCAAGTTTAAGTTCAGCAGTACAAACTGATGCTTGTAGTCCATTTTC	812
Dd	25533	NNNCCTGACTGATCCT	25592Z
Oy	813	CTTACTTTCAA	823
Dd	25593	CCTTCCTTCTTA	25603

LOCUS AX344555 349980 bp DNA linear PAT 01-FEB-2002
 DEFINITION Sequence 6 from Patent WO0200932.
 AX344555
 AX344555
 VERSION AX344555.1 GI:18492441
 KEYWORDS
 SOURCE synthetic construct.
 ORGANISM synthetic construct.
 SOURCE artificial sequences.

REFERENCE
 1
 AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
 TITLE Diagnosis of known genetic parameters within the mhc
 JOURNAL Patent: WO 0200932-A 6 03-JAN-2002;
 EpiGenomics AG (DE)
 FEATURES
 source
 1. 349980
 Location/Qualifiers
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="chemically treated genomic DNA (Homo sapiens)-Original length of seq 1: 3,673,778 <223>-split as follows.-seq 01 0.000.001 TO 0.349.980-seq 02 0.300.001 649.980-seq 03 600.001 949.980-seq 04 900.001 1.249.980-seq 05 1.200.001 1.549.980-seq 06 1.500.001 1.849.980-seq 07 1.800.001 2.149.980-seq 08 2.100.001 2.449.980-seq 09 2.400.001 2.749.980-seq 10 2.700.001 3.049.980-seq 11 3.000.001 3.349.980-seq 12 3.300.001 3.649.980-seq 13 3.600.001 3.673.778 <223>-Original length of seq 2: 3,673,778 <223>-split as follows.-seq 14 0.000.001 TO 0.349.980-seq 15 0.300.001 649.980-seq 16 600.001 949.980-seq 17 900.001 1.249.980-seq 18 1.200.001 1.549.980-seq 19 1.500.001 1.849.980-seq 20 1.800.001 2.149.980-seq 21 2.100.001 2.449.980-seq 22 2.400.001 2.749.980-seq 23 2.700.001 3.049.980-seq 24 3.000.001 3.349.980-seq 25 3.300.001 3.649.980-seq 26 3.600.001 3.673.778"

BASE COUNT 95599 a 4385 c 78257 g 171739 t
 ORIGIN

Query Match 2.2%; Score 47; DB 6; Length 349980;
 Best Local Similarity 48.0%; Pred. No. 2.5;
 Matches 134; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

QY 783 AGTACAACTGATGCTTATGCTCCATTCCTTACCTTCAATGATGATGTTGATC 842
 DB 144904 AGATGTAAAGCAATTTTGTAGTGTATTTATGTTTATGAGATTTTGTGTTT 144963
 QY 843 TTCCCTTAGCTTGTGCTTGGCGGTTTCATGAGCGTGTATGATCCCATATC 902
 DB 144964 TTGTTTATTTTGTGTTTATGTTTATTTTGTGTTTATTTTATTTT 145023
 QY 903 TGTGCTAACCTTTTACATCTACCAAGTTAAACATAGTGTATTTAGTATCTTAG 962
 DB 145024 TTTTGTAGCTTTATTTATTTAGTATTTTGTAGTGTGTATGTGAGAAATCGTTT 145083
 QY 963 TTATGCTGAGCTTATCAGGCTGTATTTGCCAACTGATGTTCAATATTTTACGTAT 1022
 DB 145084 TTTTGTAAAAATTTTGTGTTTATTTTATTTTATTTTATTTTATTTT 145143
 QY 1023 TCTTCTTAGGACCTTAATCCCTTGATCAGTTGACG 1061
 DB 145144 TTTTGTATTTATTTATTTATTTTGTGTTTGTATTTG 145182

RESULT 15
 LOCUS AX348444 7441 bp DNA linear PAT 06-FEB-2002
 DEFINITION Sequence 139 from Patent WO0202806.
 AX348444
 AX348444
 VERSION AX348444.1 GI:18614480
 KEYWORDS
 SOURCE synthetic construct.
 ORGANISM synthetic construct.
 SOURCE artificial sequences.

AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
 TITLE Method and nucleic acids for pharmacogenomic methylation analysis
 JOURNAL Patent: WO 0202806-A 139 10-JAN-2002;
 EpiGenomics AG (DE)
 FEATURES
 source
 1. 7441
 Location/Qualifiers
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 2098 a 66 c 1948 g 3329 t
 ORIGIN

Query Match 2.2%; Score 46.8; DB 6; Length 7441;
 Best Local Similarity 42.6%; Pred. No. 5.4;
 Matches 243; Conservative 0; Mismatches 327; Indels 0; Gaps 0;

QY 91 TGTGCTGCCAATTCGCTTGTGCTGCTGCAACCTGTGAATTTATCTCTTATATGAT 150
 DB 486 TGTATTTATTTTGTATTTAGTATTTATTTATTTATTTATTTATTTATTTATTT 4925
 QY 151 AAGAACTCTGCTTTTTCAGTTTTCAGTATGAGAAATTTGAGAGATCCGAGGA 210
 DB 4926 ATTTGATTTGATTTTATTTTATTTAGTATTTATTTATTTATTTATTTATTTATTT 4985
 QY 211 AGGAGACCTTTGTTTCAAGTTTGTAGCTTGTGTATTTATTTATTTATTTATTTATTT 270
 DB 4986 AAGAGAGATTTGTATTTAGATTTAGATTTTGTGAATTTGAGAGATTTTGTGTTTAT 5045
 QY 271 GTTTTCAATTTGCTATGCTTTCAGGCAACCAACCTGTTCAATCTGCAAGCC 330
 DB 5046 ATATGTTATTTGTTATTTATTTAGTTTGAAGTTTGTGAAGTTTGTGTTTATTTATTT 5105
 QY 331 TATATGATCAATTTTCAGCTTGTGAAGAGAAAGAAAGATTTCTTACTTAGAGCTCTGC 390
 DB 5106 ATATGTTATTTGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 5165
 QY 391 TGTGCTGTTAAGGAATTCGAGAGCTCTCTTTAGAAAGAAACGAGAAAGATGCTCT 450
 DB 5166 GGTGATTTGTTGATTTATTTTGTGTTTATTTTATTTATTTATTTATTTATTTATTT 5225
 QY 451 TGAACCTTCACATCTCCAGTATTTATTTATTTATTTATTTATTTATTTATTTATTT 5285
 DB 5226 TAAATTTAGAGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 5345
 QY 511 ATTTCTCATTTTCTGCTGTGATTTTGGTAAATTTGATTTGAGTTGTCTATTTAGAAC 570
 DB 5286 AATTTTCGATGATGATTTATTTGTTTATTTATTTATTTATTTATTTATTTATTTATTT 5345
 QY 571 CAACAATAATTTTACGTTTATTTAGCTGTTTATTTATTTATTTATTTATTTATTTATTT 630
 DB 5346 ATTTATTTAGATTTGTTGGAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 630
 QY 631 TCTAATCAGCAACTGTTTCAGGAAAGTA 660
 DB 5406 TTGGAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 5435

Search completed: June 24, 2003, 13:42:45
 Job time : 3733 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 18:49:36 ; Search time 31 Seconds
(without alignments)

1414.176 Million cell updates/sec

Title: US-09-870-406A-33

Perfect score: 329
Sequence: 1 MALSMERGFSGCFKAPNP.....SIIGKSYETRPMDLTGQFIG 329

Scoring table:

Capop 60.0 , Capext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	253	76.9	323	21	AA18224
2	253	76.9	327	21	AA18223
3	250	76.0	250	21	AA18225
4	123	37.4	329	23	AA50863
5	8	2.4	163	22	AA39567
6	8	2.4	263	22	AB09063
7	8	2.4	554	22	AB09133
8	8	2.4	628	22	AB09171
9	8	2.4	959	19	AA69560
10	7	2.1	11	22	AA898263

11	7	2.1	11	22	AA875183
12	7	2.1	34	21	AA158329
13	7	2.1	46	22	AB834839
14	7	2.1	46	22	AA55647
15	7	2.1	46	22	AA68027
16	7	2.1	78	21	AA834729
17	7	2.1	78	22	AA64549
18	7	2.1	79	23	AB01918
19	7	2.1	100	22	AB10160
20	7	2.1	100	23	AB25780
21	7	2.1	109	22	AA898232
22	7	2.1	109	22	AA875151
23	7	2.1	124	22	AA06786
24	7	2.1	125	22	AA89008
25	7	2.1	137	22	AA42423
26	7	2.1	142	22	AA002427
27	7	2.1	145	20	AA12361
28	7	2.1	167	22	AB099948
29	7	2.1	186	23	AB06009
30	7	2.1	194	21	AA623431
31	7	2.1	261	21	AA631121
32	7	2.1	266	22	AB852873
33	7	2.1	269	21	AA637044
34	7	2.1	276	21	AA637043
35	7	2.1	278	21	AA631120
36	7	2.1	285	22	AB625822
37	7	2.1	287	23	AB847532
38	7	2.1	294	21	AA637042
39	7	2.1	295	21	AA648028
40	7	2.1	303	21	AA659725
41	7	2.1	309	21	AA648027
42	7	2.1	310	22	AA611918
43	7	2.1	312	21	AA650418
44	7	2.1	315	21	AA631119
45	7	2.1	315	21	AA648026

ALIGNMENTS

RESULT 1	AA18224	standard; Protein: 323 AA.
ID	AA18224	
AC	AA18224	
XX		
DT	17-OCT-2000	(first entry)
XX		
DE	Arabidopsis thaliana protein fragment SRO ID NO: 19546.	
XX		
KW	Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.	
KW	Arabidopsis thaliana.	
XX		
OS	Arabidopsis thaliana.	
XX		
PN	EP1033405-A2.	
XX		
PD	06-SEP-2000.	
XX		
PF	25-FEB-2000; 2000EP-0301439.	
XX		
PR	25-FEB-1999; 99US-0121825.	
PR	05-MAR-1999; 99US-0123180.	
PR	09-MAR-1999; 99US-0123548.	
PR	23-MAR-1999; 99US-0125788.	
PR	25-MAR-1999; 99US-0126264.	
PR	29-MAR-1999; 99US-0126785.	
PR	01-APR-1999; 99US-0127462.	
PR	06-APR-1999; 99US-0128234.	
PR	08-APR-1999; 99US-0128714.	
PR	16-APR-1999; 99US-0129845.	
PR	19-APR-1999; 99US-0130077.	

A33 antigen bindin
Human Grlp-1 115A
Peptide #2345 enco
Human brain expres
Human bone marrow
Human secreted pro
Human secreted pro
Human secreted pro
Human ORF protein
Human cDNA SEQ ID
Streptococcus poly
Rabbit anti-A33 an
Rabbit anti A33 an
Human polypeptide
Human Immune/thema
Human EST encoded
Human polypeptide
Human 5' EST seque
Novel human diagno
Human ORF protein
Arabidopsis thalia
Arabidopsis thalia
Escherichia coli p
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Novel human diagno
Listeria monocytog
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Human G-protein co
Arabidopsis thalia
Arabidopsis thalia

PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131448.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143342.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149378.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154479.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0158294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.

PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 76.9%; Score 253; DB 21; Length 323;
Best Local Similarity 100.0%; Pred. No. 1,6e-242;
Matches 253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 77 YSSMTGDLGKTELOMLAFKSSKIRILRSMALNETMOVDFAGMEPEYDPIFCANFT 136
DB 71 YSSMTGDLGKTELOMLAFKSSKIRILRSMALNETMOVDFAGMEPEYDPIFCANFT 130
OY 137 STNNIVVLDLNPDLHOLDQDQDYDKYNNKMSIYHKYAEFPWGGKLTGSISIFPSFLV 196
DB 131 STNNIVVLDLNPDLHOLDQDQDYDKYNNKMSIYHKYAEFPWGGKLTGSISIFPSFLV 190
OY 197 MMTFFSSSEKHKALFSAFLFYQAMLEMTIOVREEMPSHVANRCEAOKHLYLWRAOKD 256
DB 191 MMTFFSSSEKHKALFSAFLFYQAMLEMTIOVREEMPSHVANRCEAOKHLYLWRAOKD 250
OY 257 PGHGLKRLVGEAKAKELRLDFLNGVDELGTKTFIDYFPYQEDGTGVSOKRSIIIGKSY 316
DB 251 PGHGLKRLVGEAKAKELRLDFLNGVDELGTKTFIDYFPYQEDGTGVSOKRSIIIGKSY 310
OY 317 ETRPMDLTGOFIG 329
DB 311 ETRPMDLTGOFIG 323

RESULT 2
AAG18223
ID AAG18223 standard; Protein: 327 AA.

AC AAG18223;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 19545.

KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132040.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141844.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.

PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 76.9%; Score 253; DB 21; Length 327;
Best Local Similarity 100.0%; Pred. No. 1,6e-242;
Matches 253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 YSSMTGIDGKTELOMTAFKSSKIRLRLSMAIENETMOVEDFAGFMEPEYDTPIFCANFTT 136
DB 75 YSSMTGIDGKTELOMTAFKSSKIRLRLSMAIENETMOVEDFAGFMEPEYDTPIFCANFTT 134
QY 137 STNNVIVVDLNLPHOLDTODYQDYKYYNKIMSIYHKYATFPWGKLTGESIKFSPLY 196
DB 135 STNNVIVVDLNLPHOLDTODYQDYKYYNKIMSIYHKYATFPWGKLTGESIKFSPLY 194
QY 197 MMTREFSSSKRHKALTSAPLEYQAMLEMTIOVREMEPSHVANCEAOHKYLTWRAOKD 256
DB 195 MMTREFSSSKRHKALTSAPLEYQAMLEMTIOVREMEPSHVANCEAOHKYLTWRAOKD 254
QY 257 PGHGLKRLVGEAKAKELRDPLFNGVDELGTFTFDYPEPYOTEDGTVSDDRKSIIIGKSY 316
DB 255 PGHGLKRLVGEAKAKELRDPLFNGVDELGTFTFDYPEPYOTEDGTVSDDRKSIIIGKSY 314
QY 317 ETRPMDLTGQFIG 329
DB 315 ETRPMDLTGQFIG 327

RESULT 3
AAG18225
ID AAG18225 standard; Protein; 250 AA.
XX AAG18225;
XX 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 19547.
XX
XX Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.

XX EP1033405-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135351.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148172.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.

```

PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 76.0%; Score 250; DB 21; Length 250;
 Best Local Similarity 100.0%; Pred. No. 1.2e-239;
 Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 80 MFGDGTKELOMAFASSKIRLRISMAIENETMOVDFAGFMEPEYDPTFCANFTSTN 139
DB 1 MFGDGTKELOMAFASSKIRLRISMAIENETMOVDFAGFMEPEYDPTFCANFTSTN 60
QY 140 VNIVLADLNPLOLTDQTDYODKYNKINSIYHKAETFPWGKLTGESIKFSPVMMT 199
DB 61 VNIVLADLNPLOLTDQTDYODKYNKINSIYHKAETFPWGKLTGESIKFSPVMMT 120
QY 200 RRSSEKKAHAFSALEYOAMLEMTIOVREEMESPVRANCEAOHKYLTWRAPDPGH 259
DB 121 RRSSEKKAHAFSALEYOAMLEMTIOVREEMESPVRANCEAOHKYLTWRAPDPGH 180
QY 260 GLTKRLVGEAKAKELIRDFLNGVDLGTFTFDYPEXOTEGCTVSDDRSTIGKYEYR 319
DB 181 GLTKRLVGEAKAKELIRDFLNGVDLGTFTFDYPEXOTEGCTVSDDRSTIGKYEYR 240
QY 320 PMDLTGQFTG 329
DB 241 PMDLTGQFTG 250

```

RESULT 4
 ID AAM50863 standard; Protein; 329 AA.
 AC AAM50863;
 XX

```

DT 07-MAY-2002 (first entry)
XX
DE Arabidopsis phytochromobilin synthase HY2.
XX
KW HY2; biliverdin reductase; phytochromobilin synthase;
XX phytochrome; phytofluor; plant; enzyme.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT Peptide 1..45
FT /label= Transit_peptide
FT Protein 46..329
FT /label= Mature_protein
XX
PN NC020194548-A2.
XX
PD 13-DEC-2001.
XX
PE 05-JUN-2001; 2001MO-US18326.
XX
PR 08-JUN-2000; 2000US-210286P.
PR 26-FEB-2001; 2001US-271758P.
PR 29-MAY-2001; 2001US-0210286.
XX
PA (REGC ) UNIV CALIFORNIA.
XX
PI Lagarias JC, Kochi T, Frankenberg N, Gambetta GA, Montgomery BL;
XX WPI; 2002-195566/25.
XX N-PSDB; ABA91766.
XX

```

Novel isolated HY2 family bilin reductase having bilin reductase activity, useful for converting biliverdin to phytyobilin, and for producing a photoactive holophytochrome and/or phytofluor

Example 3; Fig 3B; 102pp; English.

The present sequence is that of the HY2 protein of Arabidopsis thaliana ecotype Columbia, as predicted from cDNA sequence analysis. HY2 is a ferredoxin-dependent biliverdin reductase that has a predicted molecular mass of 38.1 kDa. The first 45 N-terminal amino acid residues are predicted to form a chloroplast transit peptide, suggesting that the HY2 protein is localised in the chloroplast. HY2 is related to a family of proteins found in oxygenic photosynthetic bacteria. It is an example of bilin reductases of the invention, which are useful e.g. for the conversion of biliverdin to phytyobilin and the assembly of holophytochromes or phytofluors. A claimed method of producing a photoactive holophytochrome involves co-expressing a haem oxygenase, an apophytochrome and a ferredoxin-dependent bilin reductase in a cell, where the cell produces the photoactive holophytochrome and where the apophytochrome and/or the bilin reductase are expressed by heterologous nucleic acids. The cell may be an algal, yeast, bacterial, plant, insect or mammalian cell, and the bilin reductase is preferably an HY2 family bilin reductase.

Sequence 329 AA;

Query Match 37.4%; Score 123; DB 23; Length 329;
 Best Local Similarity 99.6%; Pred. No. 1.8e-113;
 Matches 223; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 MALSMFEGFSGSCFAPNPVILISAPNKNFTLRKKRKLRLVSAVSYPEFSALE 60
DB 1 MALSMFEGFSGSCFAPNPVILISAPNKNFTLRKKRKLRLVSAVSYPEFSALE 60
QY 61 ETRKRIVLEPSHLOKYSMTGLDGTKELOMAFASSKIRLRISMAIENETMOVDFAG 120
DB 61 ETRKRIVLEPSHLOKYSMTGLDGTKELOMAFASSKIRLRISMAIENETMOVDFAG 120

```


QY 121 MEPEYDPICANEFTSTNNIVLNLPLHQLTDQTDYODKYYNKIMSYHKYAEPPW 180
 |||||||
 Db 121 HEPEYDPICANEFTSTNNIVLNLPLHQLTDQTDYODKYYNKIMSYHKYAEPPW 180
 |||||||
 QY 181 GKITGESIKFFSPVLMWTRPSSSKKKALFSAFLEYQAMLE 224
 |||||||
 Db 181 GKITGESIKFFSPVLMWTRPSSSKKKALFSAFLEYQAMLE 224
 |||||||

RESULT 5

AAU39567
 ID AAU39567 standard; Protein: 163 AA.

AC AAU39567;
 XX

DT 13-FEB-2002 (first entry)
 XX

DE Propionibacterium acnes immunogenic protein #463.
 XX

KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.
 XX

PN W0200181581-A2.
 XX

PD 01-NOV-2001.
 XX

PE 20-APR-2001; 2001WO-US12865.
 XX

PR 21-APR-2000; 2000US-199047P.
 XX

PR 02-JUN-2000; 2000US-208841P.
 XX

PR 07-JUL-2000; 2000US-216747P.
 XX

PA (CORI-) CORIXA CORP.
 XX

PI Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
 DR WPI: 2001-616774/71.
 PS N-PSDB: AAS59507.

PT Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris -

PS Example 1; SEQ ID No 762; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 163 AA;
 SQ

Query Match 2.4%; Score 8; DB 22; Length 163;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 303 GTVSDKRS 310
 |||||||
 Db 29 GTVSDKRS 36

RESULT 6

ABB09063
 ID ABB09063 standard; Protein: 263 AA.

AC ABB09063;
 XX

DT 26-JUN-2002 (first entry)
 XX

DE Thermus caldophilus GK24 cosmid clone 3 related protein #1.
 XX

KW Thermus caldophilus GK24; heat resistant enzyme; cosmetic; food;
 KM DNA polymerase; aqualysin I; beta-galactosidase.

OS Thermus caldophilus.
 XX

PN KR2001019888-A.
 XX

PD 15-MAR-2001.
 XX

PE 31-AUG-1999; 99KR-0036564.
 XX

PR 31-AUG-1999; 99KR-0036564.
 XX

PR (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.
 XX

PI Cho SJ, Jung BW, Kim HJ, Kim JS, Kim WC, Kim SY, Kim YS, Ko JH;
 PI Ko SH, Lee DS, Lee JS, Park JA, Park JH, Shin HJ;
 DR WPI: 2001-512441/56.

PT Base sequence and amino acid sequence of cosmid clone 3 isolated from
 PT Thermus caldophilus -

PS Example 2; Page 116; 130pp; Korean.

XX The present invention describes the nucleotide sequence of the cosmid
 CC clone 3 isolated from Thermus caldophilus GK24, as given in the present
 CC invention. Proteins encoded by the cosmid clone 3 are heat resistant
 CC enzymes which can be used in the fields of cosmetic and food industries.
 CC Thermus caldophilus GK24 has the optimal growth temperature of 72 plus
 CC degrees Celsius and produces heat resistant enzymes including DNA
 CC polymerase, aqualysin I, and beta-galactosidase. The DNA polymerase can
 CC effectively be used in PCR which is performed at high temperature. The
 CC present sequence represents a Thermus caldophilus GK24 cosmid clone 3
 CC related protein from the present invention.

XX Sequence 263 AA;
 SQ

Query Match 2.4%; Score 8; DB 22; Length 263;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 KRFLRLVS 47
 |||||||
 Db 253 KRFLRLVS 260

RESULT 7

ABB91133
 ID ABB91133 standard; Protein: 554 AA.

AC ABB91133;
 XX

DT 31-MAY-2002 (first entry)

XX DE Herbicidally active polypeptide SEQ ID NO 344.
 XX KM Herbicidal; plant; agriculture; herbicide.
 XX OS Arabidopsis thaliana.
 XX PN WO200210210-A2.
 XX PD 07-FEB-2002.
 XX PF 28-AUG-2001; 2001WO-EP09892.
 XX PR 28-AUG-2001; 2001WO-EP09892.
 XX PA (FARB) BAYER AG.
 XX PI Tietjen K, Weidler M;
 XX DR WPI; 2002-269010/31.
 XX PT Identifying plant target proteins for herbicidally active compounds,
 XX PT comprising aligning and comparing nucleic acid or amino acid sequences
 XX PT from plant with nucleic acid or amino acid sequences from non-plant
 XX PT organisms -
 XX PS Claim 5; SEQ ID NO 344; 261pp + Sequence Listing; English.
 XX CC The invention relates to identifying target proteins
 XX CC (AB90799-AB94016) for herbicidally active compounds, comprising
 XX CC aligning and comparing nucleic acid or amino acid sequences from plant
 XX CC with nucleic acid or amino acid sequences from non-plant organisms using
 XX CC suitable search parameters, where plant sequences having an E-value
 XX CC greater by a factor of 3 than the E-value of most similar non-plant
 XX CC sequences are selected. The polypeptides or nucleic acids encoding them
 XX CC are useful for identifying modulators. The identified modulators are
 XX CC useful as herbicides.
 XX SQ Sequence 554 AA;
 XX
 QY Query Match 2.4%; Score 8; DB 23; Length 554;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 274 LNRDFLFN 281
 334 LNRDFLFN 341

RESULT 8
 ABG15871
 ID ABG15871 standard; Protein; 628 AA.
 AC ABG15871;
 DT 18-FEB-2002 (first entry)
 DE Novel human diagnostic protein #15862.
 XX
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 XX Homo sapiens.
 XX OS
 XX PN WO200175067-A2.
 XX PD 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US08631.
 XX PR 31-MAR-2000; 2000US-0540217.
 XX PR 23-AUG-2000; 2000US-0649167.
 XX

PA (HYSE-) HYSEQ INC.
 XX PI Drmanac RT, Liu C, Tang YT;
 XX DR WPI; 2001-639362/73.
 XX DR N-PSDB; AAS80058.
 XX PT New isolated polynucleotide and encoded polypeptides, useful in
 XX PT diagnostics, forensics, gene mapping, identification of mutations
 XX PT responsible for genetic disorders or other traits and to assess
 XX PT biodiversity -
 XX PS Claim 20; SEQ ID NO 46230; 103pp; English.
 XX CC The invention relates to isolated polynucleotide (I) and
 XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 XX CC and gene mapping, and in recombinant production of (II). The
 XX CC polynucleotides are also used in diagnostics as expressed sequence tags
 XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
 XX CC to restore normal activity of (II) or to treat disease states involving
 XX CC (II). (II) is useful for generating antibodies against it, detecting or
 XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
 XX CC a food supplement. (II) and its binding partners are useful in medical
 XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
 XX CC disorders involving aberrant protein expression or biological activity.
 XX CC The polypeptide and polynucleotide sequences have applications in
 XX CC diagnostics, forensics, gene mapping, identification of mutations
 XX CC and to produce other types of data and products dependent on DNA and
 XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
 XX CC diagnostic amino acid sequences of the invention.
 XX CC Note: The sequence data for this patent did not appear in the printed
 XX CC specification, but was obtained in electronic format directly from WIPO
 XX CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 628 AA;
 XX
 QY Query Match 2.4%; Score 8; DB 22; Length 628;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 97 SKIRILRS 104
 595 SKIRILRS 602

RESULT 9
 AAM69560
 ID AAM69560 standard; Protein; 959 AA.
 AC AAM69560;
 DT 15-OCT-1998 (first entry)
 DE Saccharomyces cerevisiae Lig4.
 XX
 XX DNA ligase IV; Lig4; cellular DNA repair; modulation; diagnosis;
 KW XRC4; Ku-associated DNA repair pathway; cancer; retroviral infection;
 XX immune system disorder.
 XX OS
 XX Saccharomyces cerevisiae.
 XX FH Key Location/Qualifiers
 XX FT Peptide 1..15
 XX FT Protein /label= signal
 XX FT 16..959 /label= Lig4
 XX PN WO9830902-A1.
 XX PD 16-JUL-1998.
 XX

PF 13-JAN-1998; 98WO-GB00095.
XX
PR 20-JUN-1997; 97GB-0013131.
PR 13-JAN-1997; 97GB-0000574.
XX
PA (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.
XX
PI Critchlow SE, Jackson SP;
XX
DR WPI: 1998-399301/34.
DR N-PSDB: AAV40294.
XX
PT Modulation of cellular DNA repair activity - using compounds
PT identified as modulating the interaction of XRCC4, DNA ligase IV and
PT DNA-PCRs/Ku
XX
PS Example 2; Fig 6; 118pp; English.
XX
CC A method has been developed of assaying for a compound able to modulate
CC the interaction or binding between XRCC4 and any of DNA ligase IV and/or
CC DNA-PCRs/Ku (DPK). The method comprises: (a) bringing into contact: (1)
CC a substance including XRCC4 (or a fragment, derivative, variant or
CC analogue able to bind DNA ligase IV or DPK); (11) a substance including
CC DNA ligase IV (or a fragment, derivative, variant or analogue able to
CC bind XRCC4; and/or (11) DPK (or a fragment, derivative, variant or
CC analogue able to bind XRCC4), and a test compound under conditions where,
CC if the test compound is not an inhibitor of interaction or binding
CC between the substances, the substances interact or bind; and (b)
CC determining interaction or binding between the substances. Compounds
CC which modulate the interaction between XRCC4, DNA ligase IV and DPK can
CC be used to modulate cellular DNA repair activity, e.g. in the treatment
CC of proliferative disorders, cancers and tumours, disorders involving
CC retroviruses such as AIDS, human adult T-cell leukaemia/lymphoma, Type I
CC diabetes and multiple sclerosis, and also in radiotherapy and
CC chemotherapy. They can also be used in the potentiation of gene
CC targeting and gene therapy. They can also be used for the modulation of
CC immune system function. The present sequence represents Saccharomyces
CC cerevisiae Lig4 which is used in an example from the present invention.
XX
SQ Sequence 959 AA;

Query Match 2.4%; Score 8; DB 19; Length 959;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 186 GESIKFFS 193
DB 311 GESIKFFS 318

RESULT 10
AAB98263
ID AAB98263 standard; Peptide: 11 AA.
XX
AC AAB98263;
XX
DT 20-AUG-2001 (first entry)
XX
DE Anti-A33 antigen immunoglobulin VL CDRI SEQ ID NO:69.
XX
KW Human; rabbit; humanised; A33 antigen; anti-A33 antigen antibody;
KW immunoglobulin; complementarily determining region; CDR; cancer;
KW cytosolic; anticancer; colon cancer; stomach cancer.
XX
OS Homo sapiens.
XX
PN WO200130393-A2.
XX
PD 03-MAY-2001.
XX
PF 20-OCT-2000; 2000WO-US29289.
XX
PR 22-OCT-1999; 99US-0425638.
PR

PR 04-APR-2000; 2000US-0543004.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
PA (SUOK) SUOK KETTERING INST CANCER RES.
PA (SCTR) SCRIPPS RES INST.
XX
PI Barbas CF, Rader C, Ritter G, Welt S, Old LJ;
XX
DR WPI: 2001-328613/34.
XX
XX Treating cancers, particularly of stomach and colon, that express A33
PT antigen by administering conjugate of anticancer agent with specific
PT immunoglobulin product
XX
PS Claim 1; Page 36; 85pp; English.
XX
CC The present invention describes a method for treating cancers that
CC express the A33 antigen. The method comprises administering an
CC anticancer agent (I) conjugated to an immunoglobulin product (II) that
CC binds specifically to A33 and contains one or more of 13 specified
CC complementarily determining regions (CDRs), given in AAB98263 to
CC AAB98274. (I) has cytostatic activity. The method can be used for
CC treating colon and stomach cancers. (II), or the nucleic acid encoding
CC it, can be used directly, in unconjugated form, for immunotherapy of
CC cancer, and, when labeled, for detection or diagnosis of diseases
CC associated with A33 expression. AAH22218 to AAH22254 and AAB98230 to
CC AAB98371 represent sequences used in the exemplification of the
CC present invention.
XX
SQ Sequence 11 AA;

Query Match 2.1%; Score 7; DB 22; Length 11;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 277 DFLFNGV 283
DB 4 DFLFNGV 10

RESULT 11
AAB75183
ID AAB75183 standard; Peptide: 11 AA.
XX
AC AAB75183;
XX
DT 08-AUG-2001 (first entry)
XX
DE A33 antigen binding immunoglobulin product CDR peptide SEQ ID NO:69.
XX
KW Chimeric antibody; humanised; humanisation; mammalian antibody; antigen;
KW immunoreact; anti A33 antigen antibody; Immunoglobulin.
XX
OS Homo sapiens.
XX
PN WO200131065-A1.
XX
PD 03-MAY-2001.
XX
PF 20-OCT-2000; 2000WO-US29026.
XX
PR 22-OCT-1999; 99US-0425638.
PR 04-APR-2000; 2000US-0543004.
XX
PA (SCTR) SCRIPPS RES INST.
XX
PI Barbas CF, Rader C;
XX
DR WPI: 2001-328657/34.
XX
XX Preparing humanized rabbit antibodies that specifically immunoreact
PT with a particular antigen using display technology for expressing
PT libraries of antibody domains and fine tuning variable domain regions -

XX Example 9; Page 38; 62pp; English.

XX The present invention describes a method for preparing a humanised rabbit
 CC antibody that specifically immunoreacts with a particular antigen. The
 CC method comprises expressing a library of antibodies comprising one or
 CC more complementarily determining region (CDR) from the variable domain
 CC sequences that specifically immunoreact with the antigen grafted into
 CC framework regions from humans, and selecting the antibodies that react
 CC with the antigen. The method is useful for humanising non-human
 CC mammalian antibodies, which can be used for the treatment of a variety
 CC of diseases. The present sequence represents an A33 antigen binding
 CC immunoglobulin product CDR peptide which is given in an example from the
 CC present invention.

SO Sequence 11 AA;

QY Query Match 2.1%; Score 7; DB 22; Length 11;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 277 DFLFNGV 283
 4 DFLFNGV 10

RESULT 12
 AAY58329
 ID AAY58329 standard; peptide; 34 AA.

AC AAY58329;

DT 27-MAR-2000 (first entry)

XX Human Grip-1 I15A mutant NR-box 2 peptide.

DE Nuclear receptor; p160 coactivator; NR-box; coactivator binding domain;
 XX NR; AF2 transactivation domain; mutant; thyroid hormone receptor; TR;
 KW retinoid receptor; RAR; RXR; peroxisome receptor; PPAR; mulein;
 KW vitamin D receptor; VDR; oestrogen receptor; ER; glucocorticoid receptor;
 KW GR; progesterin receptor; PR; mineralocorticoid receptor; MR;
 KW androgen receptor; AR; Grip-1; Tif2; NCOA-2; RAC3; AIB1; TRAM-1; P/CIP;
 KW SRC1; breast cancer; prostate cancer; cardiac arrhythmia; infertility;
 KW osteoporosis; hyperthyroidism; hypercholesterolaemia; obesity;
 KW protein.coordinates.

XX Homo sapiens.

OS Synthetic.

XX WO960014-A2.

PN 25-NOV-1999.

PD 30-MAR-1999; 99W0-US06899.

XX 30-MAR-1998; 98US-0079956.

PR 16-DEC-1998; 98US-0113146.

XX (REGC) UNIV CALIFORNIA.

PA Baxter JD, Darimont B, Feng W, Fletcher R, Kushner PJ;
 PI Wagner RL, West BL, Yamamoto KR;
 XX WPI; 2000-072429/06.

DR Identifying modulators of nuclear receptor coactivator binding useful
 PT for generating new compounds which distinguish nuclear receptor
 PT isoforms -
 XX Disclosure; Page -: 281pp; English.

PS The invention relates a method of identification of a compound that
 CC modulates coactivator binding to a nuclear receptor. The method

CC comprises modelling test compounds that fit spatially into a nuclear
 CC receptor (NR) coactivator binding site of interest using an atomic
 CC structural model of a nuclear receptor coactivator binding site. The
 CC test compounds are screened by their ability to bind to an NR
 CC coactivator binding site, leading to the identification of a compound
 CC which modulates coactivator binding. Nuclear receptors are a superfamily
 CC of hormone/ligand activated transcription factors, and includes thyroid
 CC hormone receptors (TRs), retinoid receptors (RARs and RXRs), peroxisome
 CC receptors (PPARs and XPARs), vitamin D receptors (VDRs), oestrogen
 CC receptors (ERs), glucocorticoid receptors (GRs), progesterin receptors
 CC (PRs), mineralocorticoid receptors (MRs), androgen receptors (ARs) and
 CC iodosinoid receptors (IRs). These receptors, in addition to binding their
 CC cognate ligand, also bind coactivator proteins that are involved in
 CC receptor function, for example NRs can stimulate transcription in
 CC response to hormone binding by recruiting coactivator proteins to the
 CC promoters of responsive genes. Coactivators of the p160 family mediate
 CC activity of a transcriptional activation domain, AR2, that is part of the
 CC nuclear receptor's ligand binding domain. The method may be used to
 CC identify an agonist or antagonist of coactivator binding to a nuclear
 CC receptor. The methods can also be used to identify residues which
 CC comprise a coactivator binding site of a nuclear receptor of interest.
 CC The methods are applicable to generating new compounds that distinguish
 CC nuclear receptor isoforms. This can facilitate generation of either
 CC tissue-specific or function-specific compounds. Nuclear receptors have
 CC been implicated in a variety of medical disorders, including breast
 CC cancer, prostate cancer, cardiac arrhythmia, infertility, osteoporosis,
 CC hyperthyroidism, hypercholesterolaemia and obesity. Sequences AAY58282-
 CC AAY58302 represent the regions (NR-boxes 1-3) of p160 coactivator
 CC proteins which interact with the nuclear receptor coactivator binding
 CC sites (AAY58307-Y58328). The NR coactivator binding sites comprise two
 CC regions of the NR: helices 3-6 (H3-H6), and helix 12 (H12). The
 CC coactivator NR-boxes all share a consensus sequence motif LXXLL (AAY58279).
 CC with the NR-box 2 motifs sharing the consensus motif LXXLL (AAY58279).
 CC Sequences AAY58280-Y58281 represent the hydrophobic interaction motifs of
 CC p53 and Vp16 used in an exemplification of the invention. Sequences
 CC AAY58329-Y58240 represent mutant versions of the human Grip-1 NR-box 2
 CC sequence (AAY58283), and sequences AAY58341-Y58353 represent mutant
 CC versions of NR coactivator binding sites, used in exemplifications of the
 CC invention. Note: This sequence is not shown in the specification, but is
 CC derived from the human Grip-1 NR-box 2 sequence given in figure 6.

SO Sequence 34 AA;

QY Query Match 2.1%; Score 7; DB 21; Length 34;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 205 KEKRAL 211
 10 KEKRAL 16

RESULT 13
 ABB34839
 ID ABB34839 standard; Peptide; 46 AA.

AC ABB34839;

DT 04-FEB-2002 (first entry)

XX Peptide #2345 encoded by human foetal liver single exon probe.
 XX Human; foetal liver; gene expression; single exon nucleic acid probe.
 XX Homo sapiens.

OS WO200157277-A2.

PN 09-AUG-2001.

PD 30-JAN-2001; 2001W0-US00669.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
DR
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
XX Claim 27; SEQ ID NO 27474; 639pp + sequence listing; English.
PS
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_prt_sequences.
XX
SQ Sequence 46 AA:

Query Match 2.1%; Score 7; DB 22; Length 46;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 238 VRANCEA 244
DB 32 VRANCEA 38

RESULT 14

AAM55647
ID AAM55647 standard; Protein; 46 AA.

XX AAM55647;

XX 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe encoded protein SEQ ID NO: 27752.

KW Human; brain expressed exon; gene expression analysis; probe;
microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
epilepsy; cancer.

OS Homo sapiens.

PN WO200157275-A2.

XX 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US000667.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-483446/52.

DR WPI; 2001-483446/52.
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
XX Example 4; SEQ ID NO: 27752; 650pp + sequence listing; English.
PS
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX

SQ Sequence 46 AA:

Query Match 2.1%; Score 7; DB 22; Length 46;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 238 VRANCEA 244
DB 32 VRANCEA 38

RESULT 15

AAM68027
ID AAM68027 standard; Protein; 46 AA.

XX AAM68027;

XX 06-NOV-2001 (first entry)

DE Human bone marrow expressed probe encoded protein SEQ ID NO: 28333.

KW Human; bone marrow expressed exon; gene expression analysis; probe;
microarray; cancer; leukaemia; lymphoma; myeloma.

OS Homo sapiens.

PN WO200157276-A2.

XX 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US000668.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human bone marrow -

XX Example 4; SEQ ID NO: 28333; 658pp + sequence listing; English.

XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention.
XX

SQ Sequence 46 AA;

Query Match

Best Local Similarity 2.18; Score 7; DB 22; Length 46;
Matches 7; Conservatively 100.0%; Pred. No. 40;
Mismatched 0; Indels 0; Gaps 0;

QY 238 VRANCEA 244

DB 32 VRANCEA 38

Search completed: June 24, 2003, 18:54:25
Job time : 32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 13:42:49 ; Search time 220 Seconds
(without alignments) 14407.467 Million cell updates/sec

Title: US-09-870-406A-32

Sequence: 1 gatccccacagtcacgctg.....tcgtctctactaatcaaca 2160

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Published_Applications_NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2160	100.0	2160	US-10-159-901-32	Sequence 32, Appl
2	2160	100.0	2160	US-09-870-406A-32	Sequence 32, Appl
3	42.2	2.0	759	US-10-038-241-6	Sequence 6, Appl
4	40.2	1.9	5361	US-09-742-096-2	Sequence 2, Appl
5	40.2	1.9	6152	US-09-742-096-1	Sequence 1, Appl
6	40.2	1.9	9515	US-10-239-676-160	Sequence 160, App
7	39.6	1.8	344	US-09-960-352-1036	Sequence 1036, Ap
8	38.6	1.8	446	US-09-960-352-3400	Sequence 5785, Ap
9	38.6	1.8	516	US-09-960-352-5785	Sequence 5785, Ap
10	38.6	1.8	465237	US-09-933-267A-1	Sequence 1, Appl
11	38.2	1.8	2251	US-09-796-256A-11	Sequence 11, Appl
12	38.2	1.8	640681	US-09-790-988-1	Sequence 11, Appl
13	38	1.8	228	US-09-783-590-7986	Sequence 7986, Ap
14	38	1.8	17142	US-10-239-676-206	Sequence 206, App
15	37.8	1.7	6025	US-10-239-676-44	Sequence 44, Appl
16	37.6	1.7	404	US-09-946-807-1414	Sequence 1414, Ap
17	37.6	1.7	404	US-09-795-668-1414	Sequence 1414, Ap
18	37.6	1.7	404	US-09-795-686-1414	Sequence 1414, Ap
19	37.6	1.7	6069	US-10-239-676-172	Sequence 172, App

20	37.6	1.7	10286	9	US-10-239-676-13	Sequence 13, Appl
21	37.6	1.7	11036	9	US-10-239-676-117	Sequence 117, App
22	37.6	1.7	1503841	9	US-09-946-807-1	Sequence 1, Appl
23	37.6	1.7	1503841	10	US-09-795-668-1	Sequence 1, Appl
24	37.6	1.7	1503841	10	US-09-795-668-1	Sequence 1, Appl
25	37.6	1.7	753	10	US-09-754-634A-3	Sequence 3, Appl
26	37.2	1.7	479	9	US-09-960-352-12872	Sequence 12872, A
27	37.2	1.7	2000	9	US-09-938-842A-4483	Sequence 4483, Ap
28	37	1.7	286	10	US-09-960-352-13342	Sequence 13342, A
29	37	1.7	400	7	US-08-781-986A-1547	Sequence 1547, Ap
30	37	1.7	1691139	9	US-10-067-514-1	Sequence 1, Appl
31	36.8	1.7	12405	9	US-10-239-676-35	Sequence 122, App
32	36.6	1.7	14649	9	US-10-239-676-122	Sequence 122, App
33	36.6	1.7	40257	9	US-09-978-244A-25	Sequence 25, Appl
34	36.6	1.7	2415	9	US-09-978-385-3	Sequence 3, Appl
35	36.4	1.7	6478	9	US-10-239-676-123	Sequence 123, App
36	36.4	1.7	13049	9	US-09-966-073-1	Sequence 1, Appl
37	36.4	1.7	13049	9	US-09-872-696A-1	Sequence 1, Appl
38	36.4	1.7	13049	10	US-09-231-235-1	Sequence 1, Appl
39	36.4	1.7	13049	10	US-09-797-518A-1	Sequence 1, Appl
40	36	1.7	397	10	US-09-867-701-4679	Sequence 4679, Ap
41	36	1.7	424	9	US-09-796-692-4727	Sequence 4727, Ap
42	36	1.7	424	9	US-10-040-862-4727	Sequence 4727, Ap
43	36	1.7	730	9	US-10-198-846-2090	Sequence 2090, Ap
44	35.8	1.7	390	10	US-09-960-352-3640	Sequence 3640, Ap
45	35.8	1.7	399	10	US-09-960-352-10222	Sequence 10222, A

ALIGNMENTS

RESULT 1
US-10-159-901-32
Sequence 32, Application US/10159901
Publication No. US20030073235A1
GENERAL INFORMATION:
APPLICANT: LAGARIAS, JOHN
APPLICANT: KOICHI, TAKAYUKI
APPLICANT: FRANKENBERG, NICOLE
APPLICANT: GABRIETTA, GREGORY
APPLICANT: MONTGOMERY, BERONDA
TITLE OF INVENTION: LIGHT CONTROLLED GENE EXPRESSION UTILIZING HETEROLOGOUS PHYTOC
FILE REFERENCE: 407T-907731US
CURRENT APPLICATION NUMBER: US/10/159,901
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: 60/294,463
PRIOR FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 57
SOFTWARE: PatentIn version 3.0
SEQ ID NO 32
LENGTH: 2160
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: CDS
LOCATION: (248)..(469)
FEATURE:
NAME/KEY: CDS
LOCATION: (653)..(769)
FEATURE:
NAME/KEY: CDS
LOCATION: (852)..(947)
FEATURE:
NAME/KEY: CDS
LOCATION: (1034)..(1126)
FEATURE:
NAME/KEY: CDS
LOCATION: (1213)..(1344)
FEATURE:
NAME/KEY: CDS
LOCATION: (1419)..(1523)
FEATURE:
NAME/KEY: CDS

QY	961	AGTTATCTGGAGTTATCAGCTGTGATATGTCACAACTGATGTCCAAATTTTACTGAT	1020
Db	961	AGTTATCTGGAGTTATCAGCTGTGATATGTCACAACTGATGTCCAAATTTTACTGAT	1020
QY	1021	GTTCCTTTTAGGGACCTTAACTCCTTGATCAGTTGACTGACCCAGAGGATTTACCAAGA	1080
Db	1021	GTTCCTTTTAGGGACCTTAACTCCTTGATCAGTTGACTGACCCAGAGGATTTACCAAGA	1080
QY	1081	CAAGTATTAACAAAGATAATGTCCATATATACAAATATGTGTAGGTGACACAAAGAT	1140
Db	1081	CAAGTATTAACAAAGATAATGTCCATATATACAAATATGTGTAGGTGACACAAAGAT	1140
QY	1141	ACACCAAAATACCAATTCGCAAGTAACCTAATFGCTGAGGTGTAATGACTGATCTTGAG	1200
Db	1141	ACACCAAAATACCAATTCGCAAGTAACCTAATFGCTGAGGTGTAATGACTGATCTTGAG	1200
QY	1201	ATTTATTTTGACACTTTCCCATGGGGAGGAAATGACTGGTGAATCATTAAGTTTTC	1260
Db	1201	ATTTATTTTGACACTTTCCCATGGGGAGGAAATGACTGGTGAATCATTAAGTTTTC	1260
QY	1261	TGCGCTTTGGTGTGGAGTACGTAGTTTCGTCTACCAAGAAAAACATTAAGCTTTGTC	1320
Db	1261	TGCGCTTTGGTGTGGAGTACGTAGTTTCGTCTACCAAGAAAAACATTAAGCTTTGTC	1320
QY	1321	TCTGCGTTTCTAGAGTACTATACGATATATACTACGCGCCAAAAGCTTAAGCTTTTATG	1380
Db	1321	TCTGCGTTTCTAGAGTACTATACGATATATACTACGCGCCAAAAGCTTAAGCTTTTATG	1380
QY	1381	GAAACTTTGACGAGAGATCTATCATCTTCTCTCTACAGCATGCGCTTGAGTGCACAATCC	1440
Db	1381	GAAACTTTGACGAGAGATCTATCATCTTCTCTCTACAGCATGCGCTTGAGTGCACAATCC	1440
QY	1441	AAGTGAGGAGAGAGATGAGAACCATCTCATGTGAGAGCCAAATGTGAAGCAACACACAGT	1500
Db	1441	AAGTGAGGAGAGAGATGAGAACCATCTCATGTGAGAGCCAAATGTGAAGCAACACACAGT	1500
QY	1501	ACCTGACATGCGGACACAAAAAGTGATTTATTTCTTTTGTGTAAATTTGCATGTTTGA	1560
Db	1501	ACCTGACATGCGGACACAAAAAGTGATTTATTTCTTTTGTGTAAATTTGCATGTTTGA	1560
QY	1561	ACAGACACTGATCTGATTTCTTACAAATGATATTGATTTGGTGTTCGACGATNCCGGA	1620
Db	1561	ACAGACACTGATCTGATTTCTTACAAATGATATTGATTTGGTGTTCGACGATNCCGGA	1620
QY	1621	CATGGTCTCTCTTAAAAAGTTAGTGGTGAAGCAAAAGGATTAATAAAGATTGATCC	1680
Db	1621	CATGGTCTCTCTTAAAAAGTTAGTGGTGAAGCAAAAGGATTAATAAAGATTGATCC	1680
QY	1681	CATTAGTGTCCCATTTATTAATTAAGTGTGAGATTTTGAANAATGTTGAACAAATTC	1740
Db	1681	CATTAGTGTCCCATTTATTAATTAAGTGTGAGATTTTGAANAATGTTGAACAAATTC	1740
QY	1741	AGGAGCTGTAAAGGATTTCTCTGTCAATGGGTGGATGACTTTAGCCACAAAAACATTCA	1800
Db	1741	AGGAGCTGTAAAGGATTTCTCTGTCAATGGGTGGATGACTTTAGCCACAAAAACATTCA	1800
QY	1801	TTGATTACTTCCAGATACCAAAACAGAAATGAACTGTAAGCCGTTAAACGAAGATCA	1860
Db	1801	TTGATTACTTCCAGATACCAAAACAGAAATGAACTGTAAGCCGTTAAACGAAGATCA	1860
QY	1861	TTGGGAGTCAATGAAGTCTGTCATGGATTTTAAACAGACAAATTAATCGGCTAACAT	1920
Db	1861	TTGGGAGTCAATGAAGTCTGTCATGGATTTTAAACAGACAAATTAATCGGCTAACAT	1920
QY	1921	GATTAATGTAACACAGTACATTTCCAGAGTCACATCAACAAGAGACGTAAGCTTAAGGA	1980
Db	1921	GATTAATGTAACACAGTACATTTCCAGAGTCACATCAACAAGAGACGTAAGCTTAAGGA	1980
QY	1981	AGTAGAATAAAGAAAGACAGCATGAGAGTCTCTACGCTCTATCTGATTTCAAGATGA	2040
Db	1981	AGTAGAATAAAGAAAGACAGCATGAGAGTCTCTACGCTCTATCTGATTTCAAGATGA	2040

QY 2041 TTGTTGAGTACATCATGATTTAGTTTACAGAGTGTAGCTCAGCCCTTCATCAAAAT 2100
 Db 2041 TTGTTGAGTACATCATGATTTAGTTTACAGAGTGTAGCTCAGCCCTTCATCAAAAT 2100
 QY 2101 GAGATCTCTGAGTATATATGATTTTAAAGAAATATATCTCTCTACCTAATCAACA 2160
 Db 2101 GAGATCTCTGAGTATATATGATTTTAAAGAAATATATCTCTCTACCTAATCAACA 2160

RESULT 2

US-09-870-406a-32
 ; Sequence 32, Application US/09870406A
 ; Publication No. US20030104379A1
 ; GENERAL INFORMATION:
 ; APPLICANT: LAGARIAS, JOHN
 ; APPLICANT: KOICHI, TAKAYUKI
 ; APPLICANT: FRANKENBERG, NICOLE
 ; APPLICANT: GAMBETTA, GREGORY
 ; APPLICANT: MONTGOMERY, BERONDA
 ; TITLE OF INVENTION: HY2 FAMILY OF BILIN REDUCTASES
 ; FILE REFERENCE: 407T-907720US
 ; CURRENT APPLICATION NUMBER: US/09/870,406A
 ; CURRENT FILING DATE: 2002-09-04
 ; PRIOR APPLICATION NUMBER: 60/271,758
 ; PRIOR FILING DATE: 2001-02-26
 ; PRIOR APPLICATION NUMBER: 60/210,286
 ; PRIOR FILING DATE: 2000-06-08
 ; NUMBER OF SEQ ID NOS: 57
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 32
 ; LENGTH: 2160
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (248)..(469)
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (653)..(769)
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (852)..(947)
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1034)..(1126)
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1213)..(1344)
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1419)..(1523)
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1612)..(1662)
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1743)..(1913)
 ; US-09-870-406a-32

Query Match 100.0%; Score 2160; DB 9; Length 2160;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2160; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCCTCCACGTCACGTCGATTCGATTCACGTCGCGGATGTGGGCCCTATAGTTGG 60
 Db 1 GAATTCCTCCACGTCACGTCGATTCGATTCACGTCGCGGATGTGGGCCCTATAGTTGG 60
 QY 61 ACCATGACTCGAGAGATGTGAATTCATTTGCTGCCAATTCGCTTGTCTCAGTGA 120
 Db 61 ACCATGACTCGAGAGATGTGAATTCATTTGCTGCCAATTCGCTTGTCTCAGTGA 120
 QY 121 AACTGTAATAATTTATCTCTTTATAGATTAAGAAATCTGCTTTTTCAGTTTCAGTA 180
 Db 121 AACTGTAATAATTTATCTCTTTATAGATTAAGAAATCTGCTTTTTCAGTTTCAGTA 180

Db 121 AACTGTAATAATTTATCTCTTTATAGATTAAGAAATCTGCTTTTTCAGTTTCAGTA 180
 QY 181 TGAAGAAANTTGAAGAGATGTCCGAGAGAGACCTTGTGTTAGTTTGAGACTCT 240
 Db 181 TGAAGAAANTTGAAGAGATGTCCGAGAGAGACCTTGTGTTAGTTTGAGACTCT 240
 QY 241 TGTGTATGTGCTTATCAATGAGTTGGGTTTTCATATGGGTCATGCTTCAAGGCACC 300
 Db 241 TGTGTATGTGCTTATCAATGAGTTGGGTTTTCATATGGGTCATGCTTCAAGGCACC 300
 QY 301 AAACCCACCTGTTCTAACTCTGCAAGCCCTTAATAGATCAATTCACGTTGAGAGAG 360
 Db 301 AAACCCACCTGTTCTAACTCTGCAAGCCCTTAATAGATCAATTCACGTTGAGAGAG 360
 QY 361 AAAGAAAAGATCTTACTTAGAGTCTGCTGCTGCTGCTATTAAGAAATTCGAGAGTGC 420
 Db 361 AAAGAAAAGATCTTACTTAGAGTCTGCTGCTGCTGCTATTAAGAAATTCGAGAGTGC 420
 QY 421 TTGAGAGAAACGAGAAAGAGATGCTTCTGAACCTTCACATCTCCAGATATGCAAT 480
 Db 421 TTGAGAGAAACGAGAAAGAGATGCTTCTGAACCTTCACATCTCCAGATATGCAAT 480
 QY 481 TACATTTGCTTAGTGTAGTGGAGATTAATTTCTCATGTTCTTCTGTGAATTTTG 540
 Db 481 TACATTTGCTTAGTGTAGTGGAGATTAATTTCTCATGTTCTTCTGTGAATTTTG 540
 QY 541 GGTAAATGATTAAGTTGTCATAGAACCAAACTTACTGTTATGACTGC 600
 Db 541 GGTAAATGATTAAGTTGTCATAGAACCAAACTTACTGTTATGACTGC 600
 QY 601 TTATATAGTAAAGTTTCAGATTTTGTCTTCTAATCAGCAAACTGTTTCGAGAAAGTA 660
 Db 601 TTATATAGTAAAGTTTCAGATTTTGTCTTCTAATCAGCAAACTGTTTCGAGAAAGTA 660
 QY 661 TAGTAGCATGACAGACTAGATGTAAGACCGAACTTCAATGCTTCTTAAATCTTC 720
 Db 661 TAGTAGCATGACAGACTAGATGTAAGACCGAACTTCAATGCTTCTTAAATCTTC 720
 QY 721 AAAGATTAAGCTCTGAGGAGATGGAATGAAATGAGACATGAGGTTTAACCTCA 780
 Db 721 AAAGATTAAGCTCTGAGGAGATGGAATGAAATGAGACATGAGGTTTAACCTCA 780
 QY 781 GCAGTACAAACTGATGCTTGTAGTCCATTTCTTACTTCAATGATGATGTTTGTGA 840
 Db 781 GCAGTACAAACTGATGCTTGTAGTCCATTTCTTACTTCAATGATGATGTTTGTGA 840
 QY 841 TCTTCGCTTAGTCTTGTGACTTTCGCGGTTTCATGAGCCGATGATGATCTCCATAT 900
 Db 841 TCTTCGCTTAGTCTTGTGACTTTCGCGGTTTCATGAGCCGATGATGATCTCCATAT 900
 QY 901 TCTGTGCTAACTTTTACATCTACCAACGTTAATGATGATGATGATGATGATGAT 960
 Db 901 TCTGTGCTAACTTTTACATCTACCAACGTTAATGATGATGATGATGATGATGAT 960
 QY 961 AGTATGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 Db 961 AGTATGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 QY 1021 GTTCTCTTAGGAGCCTTAATCTTGCATGATGATGATGATGATGATGATGATGATGAT 1080
 Db 1021 GTTCTCTTAGGAGCCTTAATCTTGCATGATGATGATGATGATGATGATGATGATGAT 1080
 QY 1081 CAAGTATTAATTAACAAGATTAATGTCATATATACAAATTAATGCTGAGTACCAAGAT 1140
 Db 1081 CAAGTATTAATTAACAAGATTAATGTCATATATACAAATTAATGCTGAGTACCAAGAT 1140
 QY 1141 ACACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1200
 Db 1141 ACACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1200
 QY 1201 ATTTATTTGACAGACTTTTCCATGAGGAGGAAATGATGATGATGATGATGATGATGAT 1260
 Db 1201 ATTTATTTGACAGACTTTTCCATGAGGAGGAAATGATGATGATGATGATGATGATGAT 1260

RESULT 3
US-10-038-241-6
Sequence 6, Application US/10038241
Publication NO US20030032167A1
GENERAL INFORMATION:
APPLICANT: Conklin, Darrell C.
APPLICANT: Gao, Zeren
APPLICANT: Lofton-Day, Catherine E.
APPLICANT: Whitmore, Theodore E.
TITLE OF INVENTION: SECRETED ALPHA-HELICAL PROTEIN ZLMDA24
FILE REFERENCE: 00-94

Query Match	2.08;	Score 42.2;	DB 9;	Length 759;
Best Local Similarity	26.48;	Pred. No. 1.2;		
Matches 98;	Conservative 62;	Mismatches 211;	Indels 0;	Gaps 0;

RESULT 4
 US-09-742-096-2/c
 Sequence 2, Application US/09742096
 Patient No. US20020155441A1
 GENERAL INFORMATION:
 APPLICANT: DROULHE, PIERRE
 APPLICANT: DAUBERSIES, PIERRE
 TITLE OF INVENTION: MALAIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
 FILE REFERENCE: 200773USODIV
 CURRENT APPLICATION NUMBER: US/09/742, 096
 CURRENT FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: US 08/973,642
 PRIOR FILING DATE: 1998-02-06
 PRIOR APPLICATION NUMBER: PCT/FR96/00894
 PRIOR FILING DATE: 1996-06-12
 PRIOR APPLICATION NUMBER: FR 95/07007
 PRIOR FILING DATE: 1995-06-13
 NUMBER OF SEQ ID NOS: 29
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO: 2
 LENGTH: 5361
 TYPE: DNA


```

; LENGTH: 344
;
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 05-LIB3058-031-01-K1-B1
US-09-960-352-1036

```

Query Match	1.8%;	Score 39.6;	DB 10;	Length 344;
Best Local Similarity	46.9%;	Pred. No. 3.5;		
Matches 123; Conservative	0;	Mismatches 139;	Indels 0;	Gaps 0;

QY	799	TTTGGTCCATCTCTTACCTTCATTAATTCATTCATTCGCTAGGCTTTG	858
Db	9	TTTTTTGTTTTTTTTTTTTTATNCATTTATTTATTTTTTCTGTGGCTTTTTTTTTT	68
QY	859	ACTTTGGGGTTTCATGGAGCCCTGAGTAGTACTCCATATCTGTCTAATTTTCA	918
Db	69	TATTTTATCTTTTTTTTATTTATTTTTTTTATTTTTTTTATTTTTTTTTTTTTTTT	128
QY	919	CATTCACCAAGCTTAACATAGTATGATTTGTAAGTATCTTCAGTATAGCGAGTATAC	978
Db	129	TTTTTTTTTTCTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTATTTTTTTTTTAAATTT	188
QY	979	AGGTCCTGATTTGCCAAACGATGTCATATTTTACTGATAGTCTCTTTAGGAGACT	1038
Db	189	TTTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT	248
QY	1039	TAACTCTTGCAATCAGTTCAGCT	1060
Db	249	TTTTTTTTTTTTTATATATTT	270

```

RESULT 8
US-09-960-352-3400
: Sequence 3400, Application US/09960352
: Patent No. US20020137139A1
: GENERAL INFORMATION:
: APPLICANT: Warren, Wesley C.
: APPLICANT: Tao, Nengding
: APPLICANT: Byatt, John C.
: APPLICANT: Mathialagan, Nagappan
: TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION ANT
: FILE REFERENCE: 16511.006/37-21.102987C
: CURRENT APPLICATION NUMBER: US/09/960,352
: CURRENT FILING DATE: 2001-09-24
: NUMBER OF SEQ ID NOS: 15112
: SEQ ID NO 3400
: LENGTH: 446
: TYPE: DNA
: ORGANISM: Bos taurus
: OTHER INFORMATION: Clone ID: 15-LIB3058-052-Q1-K1-D11
US-09-960-352-3400

```

Query Match	1.8%;	Score 38.6;	DB 10;	Length 446;
Best Local Similarity	47.0%;	Pred. No. 7.5;		
Matches 119; Conservative	0;	Mismatches 134;	Indels 0;	Gaps 0

QY	795	TTCGTTAGGCCAATTCCTACTTCAATTGAATGGATGGTAGTGTAACCTCCGTAGATC	854
Db	96	TT	155
QY	855	TTTGACATTTCGGGTTTCATGAGACCCTGAGATGATACTCCCATATCTGTGCACTTT	914
Db	156	TT	215
QY	915	TTGCACATCTCCCAACGTTAACATAGTTGATTTGAAGTAACTTCTCTAGTAGCTGAGT	974
Db	216	TTTTTTTTTTTAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTAAATTTTTTTTTTTAAAT	275
QY	975	TATCAGGCTCGTATATGTCCAAACGATGTCAAATATTACTGATAGTCTCTTTAGG	1033
Db	276	TTTTTATTTAATTAATTTATTTTTTTTTTAAATTTTAAATTTTTTTTTTTTTTTTTTTT	335

QY	1035	ACCTTAATCC	TTT	1047
Db	336	TTTTTTATTT	TTTT	348

```

RESULT 9
US-09-960-352-5785
: Sequence 5785, Application US/09960352
: Patent No. US20020137139A1
: GENERAL INFORMATION:
: APPLICANT: Warren, Wesley C.
: APPLICANT: Tao, Nengping
: APPLICANT: Byatt, John C.
: APPLICANT: Mathalagan, Nagappan
: TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
: TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
: FILE REFERENCE: 16511.006/37-21(10298)C
: CURRENT APPLICATION NUMBER: US/09/960,352
: CURRENT FILING DATE: 2001-09-24
: NUMBER OF SEQ ID NOS: 15112
: SEQ ID NO 5785
: LENGTH: 516
: TYPE: DNA
: ORGANISM: Bos taurus
: FEATURE:
: NAME/KEY: unsure
: LOCATION: (76),(90)
: OTHER INFORMATION: unsure at all n locations
: OTHER INFORMATION: Clone ID: 25-LIB3057-018-Q1-K1-G1
US-09-960-352-5785

```

Query Match	1.88;	Score 38.6;	DB 10;	Length 516;
Best Local Similarity	47.7%;	Pred. No. 8.2;		
Matches 113; Conservative	0;	Mismatches 124;	Indels 0;	Gaps 0;

QY	795	TTGCTTTAGTCCCACTTCTCCACTTCCAAATGGATGATGGTTGATCTTCGCTCAAGGC	854
Db	101	TGTGATTAATTTTATTTTTTTTTTTTTTTTATTTTTTTTATTTTTTTTTTTTTTTTTTTA	160
QY	855	TTTGACCTTTCGGGGTTTCACGAGCCCTGAGTAGATACCTCCCAATTCGTGCACATT	914
Db	161	TATTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTAAATTTTTTT	220
QY	915	TTGCATCTACCAACGTTTACATAGTGGTATGTAAGTATCTCTAGTATGCGGAGT	974
Db	221	TTATTTTTTTTTTTTTTAAATTTTTTAAATTTTTTTTTTATTTTTTTTTTAAATTTTAAAT	280
QY	975	TATCAGGTGCTGATGTCACAAACGANGTCCAAATTTTACTGATGCTCTCTTA	1031
Db	281	TTTTTAATTTTTTTTTTTTTTTTATTTAATATTTTTTTTTTAAATTTTTTTTTTTTTTTA	337

RESULT 10
US-09-933-267A-1
Sequence 1, Application US/099333267A
Patient No. US20020123095A1
GENERAL INFORMATION:
APPLICANT: Kalush, Francis et al.
TITLE OF INVENTION: Estrogen receptor alpha variants and
FILE OF INVENTION: methods of detection thereof
FILE REFERENCE: CU000258614
CURRENT APPLICATION NUMBER: US/09/933,267A
CURRENT FILING DATE: 2001-08-21
PRIOR APPLICATION NUMBER: 60/160626
PRIOR FILING DATE: 1999-10-20
PRIOR APPLICATION NUMBER: 60/183756
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: 09/652414
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 09/768184
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: 09/804076
PRIOR FILING DATE: 2001-03-13

PRIOR APPLICATION NUMBER: 09/826314
PRIOR FILING DATE: 2001-04-05
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 465237
TYPE: DNA
ORGANISM: human
US-09-933-267A-1

Query Match 1.8%; Score 38.6; DB 10; Length 465237;
Best Local Similarity 49.7%; Pred. No. 4e+02;
Matches 98; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 473 TATGCAATTTACATTCCTTACGTCAGTGGAGATTAATTTTCATGTTCTTCTGCT 532
DB 309882 TATGATCTACTCTTTTATATTTATATGAGAAAGCTTTCCCTTATGTTGCTCT 309941
QY 533 GAATTTGGTAATGATTTGAGTGTCAATTAGAACCAAAATTAATCTTACTGTTA 592
DB 309942 GAATTTGTTATATTTAATAGAAATTTATAGCCTATCCCTGGTAAATTCAGTTT 310001
QY 593 TAGACGCTTATATTAAGTAAAGTTCAAGATTTGTTTCTTATCAGCAAACTGTTTCA 652
DB 310002 TTTTAAAGTAGAATAATTAATCTTATTTACTATTTTACATGATGTAACCTAGTC 310061
QY 653 GAAAGTATAGTAGCAT 669
DB 310062 AATTAATGTAATATCAT 310078

RESULT 11
US-09-796-256A-11/c
Sequence 11, Application US/09796256A
Patent No. US20020078477A1
GENERAL INFORMATION:
APPLICANT: Chiang, Vincent L
APPLICANT: Carraway, Daniel H
APPLICANT: Smeltzer, Richard H
TITLE OF INVENTION: Production of Syringyl Lignin in Gymnosperms
FILE REFERENCE: 50617/c-3532.0
CURRENT APPLICATION NUMBER: US/09/796,256A
CURRENT FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US60/033381
PRIOR FILING DATE: 1996-12-16
PRIOR APPLICATION NUMBER: 08/991677
PRIOR FILING DATE: 1997-12-12
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 11
LENGTH: 2251
TYPE: DNA
ORGANISM: Pinus taeda
US-09-796-256A-11

Query Match 1.8%; Score 38.2; DB 10; Length 2251;
Best Local Similarity 49.7%; Pred. No. 24;
Matches 97; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 1023 TCTTCTTTAGGACCTTAATCTTTCATAGTGAAGTGAACCAAGGATTCACAGACA 1082
DB 560 TATATTTTAAACCTCAATACATACAAATAAATTAATAAATAAATCCCAAAATCA 501
QY 1083 AGATTTTAAACAGATATGCTCATATATGACAAATATGCGAGAGTACCCCAAGATAC 1142
DB 500 CTATTTAAACCCCTTATTCATATAAATAAAGATCGTTAATGATTCGAATCTAA 441
QY 1143 ACCAATTTACTCAATTCAGTAAGTAACTTAATGCTGAGGTAAATGACTGATCTGAGAT 1202
DB 440 AACATAAGATTAAGACAGACAACTCATTTTAAAGGTAAAGTTTCAAGTTTGT 381
QY 1203 TTAATTCAGACTTT 1217

DB 380 ATGTTTCAAAATTT 366

RESULT 12
US-09-790-988-1/c
Sequence 1, Application US/09790988
Patent No. US20020127687A1
GENERAL INFORMATION:
APPLICANT: SHIGENOBU, SHUJI
APPLICANT: WATANABE, HIDEKI
APPLICANT: HATTORI, MASAHIRA
APPLICANT: SAKAKI, YOSHIYUKI
TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
FILE REFERENCE: 081356/0159
CURRENT APPLICATION NUMBER: US/09/790,988
CURRENT FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: JP2000-107160
PRIOR FILING DATE: 2000-04-07
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 640681
TYPE: DNA
ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 1.8%; Score 38.2; DB 10; Length 640681;
Best Local Similarity 45.9%; Pred. No. 5.6e+02;
Matches 130; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 748 AATGAGATGAGCAATGACAGGTTTACTTACAGACAGACAACTGATTTGCTTAGGCC 807
DB 260596 AATAAAATATACGATTTCTTTTTCCTCTCTGTTCAATTAATCAATCAGGAGA 260537
QY 808 ATTTCTTACTTCAATGATGATGTTGTTGATCTTCGCTTAGGCTTTGACCTTGGCG 867
DB 260536 AGTTCACGAGATGATTAATGAGAAATGATTAATTTTTCGTAATTTTATACATC 260477
QY 868 GTTTCATGAGACCTGAGATATATCTCCATATTTCTGCTGCTTATCTTTTCATCTACA 927
DB 260476 ATTCCTCTCTGCTGCTGATTAATATTTTAAACGCTTCAATTAATGCTCCATCTCCA 260417
QY 928 ACGTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 987
DB 260416 CCATTAATTTGCTGCTATATTTAAATTAACCGGATTTGCTCAAAAGTATTTGTTCTGCA 260357
QY 988 TTGTCAAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1030
DB 260356 TATGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 260314

RESULT 13
US-09-783-590-7986/c
Sequence 7986, Application US/09783590
Patent No. US20020110850A1
GENERAL INFORMATION:
APPLICANT: Dillon, Patrick J.
APPLICANT: Haseltine, William A.
APPLICANT: Li, Haodong
APPLICANT: Ruben, Craig A.
APPLICANT: Rosen, Steven M.
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
FILE REFERENCE: PO-16,2C1
CURRENT APPLICATION NUMBER: US/09/783,590
CURRENT FILING DATE: 2000-02-15
PRIOR APPLICATION NUMBER: 08/420,856
PRIOR FILING DATE: 1995-04-12
PRIOR APPLICATION NUMBER: 08/346,731
PRIOR FILING DATE: 1994-11-21
NUMBER OF SEQ ID NOS: 12485
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7986
LENGTH: 228

Query Match	1.83;	Score 38;	DB 9;	Length 17142;
Best Local Similarity	49.58;	Pred. No. 89;		
Matches 98;	Conservative	0;	Mismatches 100;	Indels 0;
			Gaps	0

QY	502	AAGGATTAATATTTCTCATGTTCTTCGTGGTAATTTGGGTAATTAATGAGTGC	561
Db	2763	GTGGTTACGTTTGTAATTAAGTTTTCGGAGGTTAAGCGGGCGGATTAATTGAGGT	2822
QY	552	ATTAGGAACCAACAATACTACTCTATAGACTGCTAATATAGTAAAGTTACGA	621
Db	2823	TAGAGATTCGAGATTAGTTGATTAAATAGTGGAATTTGTTTATATAAAAATATAA	2882
QY	622	TTTCTTTTCTAATCAGAAACGTTACAGAAAGATATGACATGACAGAGCTGA	681
Db	2883	AATTGTTAGGATATAGTGATTTGTTGTACTTTTACTATTAGAGAGGTGAGGTGAG	2942
QY	682	TGGTAGACCGAATCTCAATGCTGCTTTAAATCTTCAAGATTAGCTGTGAGAG	741
Db	2943	AGAATTATTTGAAATTTGGAGGTGAGGTGATGAGATTTGATATGTTATTT	3002
QY	742	TATGCAATAGAGATGAGACATGACAGTTAACTTACGACAGTCAAACTGATGCTTT	801
Db	3003	TAGTTTGGCCGTAATAGTGAGTTTTATTTTTTAAAAAATAAAAATGATTAATAACG	3062

```
QY      802 AGGCCATTTCCTTCAATTGATTGATTGTTGATCTGCGCTTAGGCTTTGACT 861
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      3063 GATTAAGTGGATTTTATTATAGAGATTATTATGTTATTTAGATTATTTAAATAATT 3122
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      862 TTGCGGTTTCATGAGAGCCGAGTATGATACCTCCCATATTCTGTCGCTAACTTTTCACAT 921
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      3123 TTGTTGATCGGTTGTTATTTTGTGTTAGAGTGTGTAGTTATTTTGTGAA 3182
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      922 CTACCAAGTTAATAGTGTGTTAGTTAGTTACTCTCTAGTTATGCTGAGTTATCAG 981
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      3183 GAGAAAAGATAGTATAGATGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 3242
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      982 TCTGTATTGCCAAACTGATGTCATATTTTACTGATGTTCTTT 1030
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      3243 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 3291
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: June 24, 2003, 15:35:02
Job time : 227 secs

Signal transductio
Lycopodium penn
Lycopodium penn
Human immune sy
Human immune sy
Human immune sy
Human immune sy
Chemically treat
Tumour suppressor
Human immune sy
Teast AOD9604-
Human immune sy
Human immune sy
Human immune sy
Human metastasis
Chemically treated
Signal transductio
Chemically treated
Signal transductio
Human immune sy
Chemically treated
Human gene regulat
Human angiogenesis
Human immune sy
Human immune sy
Human immune sy
Oligonucleotide fo
Oligonucleotide fo
Tumour suppressor
Human immune sy
Human immune sy
Human metastasis
Mouse zlmada2 dege
Tumour suppressor
Arabidopsis thalian
Human immune sy
Human immune sy
Drosophila melanog

FT	intron	770..851
FT		/*tag= f
FT		/number= 2
FT	exon	852..946
FT		/*tag= g
FT		/number= 3
FT	intron	947..1032
FT		/*tag= h
FT		/number= 3
FT	exon	1033..1126
FT		/*tag= i
FT	intron	/number= 4
FT		1127..1212
FT		/*tag= j
FT	exon	/number= 4
FT		1213..1344
FT		/*tag= k
FT	intron	/number= 5
FT		1345..1418
FT		/*tag= l
FT		/number= 5
FT	exon	1419..1523
FT		/*tag= m
FT	intron	/number= 6
FT		1524..1611
FT		/*tag= n
FT	exon	/number= 6
FT		1612..1662
FT		/*tag= o
FT	intron	/number= 7
FT		1663..1742
FT		/*tag= p
FT	exon	/number= 7
FT		1742..2148
FT		/*tag= q
FT	variation	/number= 8
FT		replace(234,TT)
FT		/*tag= r
FT		/standard_name= "Single nucleotide polymorphism"
FT		/note= "inserted T at nucleotide 234 in Landsberg erecta and Wassilewskija ecotypes"
FT	variation	replace(364,T)
FT		/*tag= s
FT		/standard_name= "Single nucleotide polymorphism"
FT		/note= "amino acid change to Asn in Landsberg erecta and Wassilewskija ecotypes"
FT	variation	replace(515,A)
FT		/*tag= t
FT		/standard_name= "Single nucleotide polymorphism"
FT		/note= "in intron, Landsberg erecta ecotype"
FT	variation	replace(844,A)
FT		/*tag= u
FT		/standard_name= "Single nucleotide polymorphism"
FT		/note= "silent, Landsberg erecta ecotype"
FT	variation	replace(1145,T)
FT		/*tag= v
FT		/standard_name= "Single nucleotide polymorphism"
FT		/note= "in intron, Landsberg erecta ecotype"
FT	variation	replace(1910,T)
FT		/*tag= w
FT		/standard_name= "Single nucleotide polymorphism"
FT		/note= "silent, Wassilewskija ecotype"
PN		
PN	MO200194548-A2.	
XX		
XX	13-DEC-2001.	
PD		
PE	05-JUN-2001; 2001MO-US18326.	
XX		
XX	08-JUN-2000; 2000US-210286F.	
PR	26-FEB-2001; 2001US-271758P.	
PR	29-MAY-2001; 2001US-0210286.	
XX		

PA (REGC) UNIV CALIFORNIA.
XX
PI Lagarias JC, Kochi T, Frankenberg N, Gambetta GA, Montgomery BL,
XX
XX WPI: 2002-195566/25.
DR P-PSDB; AAM50863.
DR
XX
XX Novel isolated HY2 family bilin reductase having bilin reductase
PT activity, useful for converting biliverdin to phytylbilin, and for
PT producing a photoactive holophytochrome and/or phytylfur -
XX
XX Example 3; Fig 3B; 102pp: English.
XX

The present sequence is that of the HY2 gene of *Arabidopsis thaliana* ecotype Columbia on chromosome 3, with single nucleotide polymorphisms found in ecotypes Landsberg erecta and Wassilewskaja indicated. A positional cloning strategy was used to isolate the HY2 gene. For fine mapping, an HY2-1 mutant of ecotype Landsberg erecta was outcrossed with wild-type ecotype Columbia, and a mapping population was selected from F2 families with a long hypocotyl phenotype. This localised the gene to 2 overlapping bacterial artificial chromosome clones, and the gene was finally identified by sequencing of wild-type and mutant alleles. The HY2 gene encodes a ferredoxin-dependent biliverdin reductase, phytylchromobilin synthase (see AAM50863), that is related to a family of proteins found in oxygenic photosynthetic bacteria. The HY2 gene product is localised in the chloroplast. HY2 is an example of bilin reductases of the invention, which are useful e.g. for the conversion of biliverdin to phytylbilin and the assembly of holophytochromes or phytofluors. The cloning of the *Arabidopsis* HY2 gene will help to identify phytylchromobilin synthase genes from other plant species. A claimed method of producing a photoactive holophytochrome involves co-expressing a haem oxygenase, an apophytochrome and a ferredoxin-dependent bilin reductase in a cell, where the cell produces the photoactive holophytochrome and where the apophytochrome and/or the bilin reductase are expressed by heterologous nucleic acids. The cell may be an algal, yeast, bacterial, plant, insect or mammalian cell, and the bilin reductase is preferably an HY2 family bilin reductase.

SQ Sequence 2160 BP; 645 A; 361 C; 461 G; 693 T; 0 other;

Query Match	100.0%	Score 2160;	DB 24;	Length 2160;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 2160; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	1	GAATTC	CCCCACGCTCAACGTGACTGTGCATTCACAGCTGGCGGATGTGGCCCTAATAGTTGG	60
Db	1	GAATTC	CCCCACGCTCAACGTGACTGTGCATTCACAGCTGGCGGATGTGGCCCTAATAGTTGG	60
QY	61	ACCATGACTCGGACGGATGTTGAAATTCATTGTGTCGCCAATTCGCTTTGTCTCACTGA	120	
Db	61	ACCATGACTCGGACGGATGTTGAAATTCATTGTGTCGCCAATTCGCTTTGTCTCACTGA	120	
QY	121	AACCTGGAATTTTATCTCTTTTATAGATTAAGAATCTTGCCTTTTTCAGTTTTCAGTA	180	
Db	121	AACCTGGAATTTTATCTCTTTTATAGATTAAGAATCTTGCCTTTTTCAGTTTTCAGTA	180	
QY	181	TGAAGAAATTTGAAGAGAGTGTCCGAGGAAGAGACCTTGGTTTCACTTTGTGTAGCT	240	
Db	181	TGAAGAAATTTGAAGAGAGTGTCCGAGGAAGAGACCTTGGTTTCACTTTGTGTAGCT	240	
QY	241	TGTTGTAATGGCTTTATCAATGAGAGTTGGGTTTCAATTGGGTCAATGCTTTAAGGCACC	300	
Db	241	TGTTGTAATGGCTTTATCAATGAGAGTTGGGTTTCAATTGGGTCAATGCTTTAAGGCACC	300	
QY	301	AAACCCACCTGTTCTAATCTCTGSCAAGCCCTAATAAGATCAATTCACGTTGAGAAAGAG	360	
Db	301	AAACCCACCTGTTCTAATCTCTGSCAAGCCCTAATAAGATCAATTCACGTTGAGAAAGAG	360	
QY	361	AAAGAAAAGATCTTACTTGAAGTCTCTGCTGTGTCGTAATAGGAATTCGACAGAGTGGC	420	
Db	361	AAAGAAAAGATCTTACTTGAAGTCTCTGCTGTGTCGTAATAGGAATTCGACAGAGTGGC	420	

421 TTTGAGAAACCGAGAAAGATCGTTCTTGAACTTCACATCTCCAGATATGCAAT 480
421 TTTGAGAAACCGAGAAAGATCGTTCTTGAACTTCACATCTCCAGATATGCAAT 480
481 TACATTCGTAGTGTAGTGGAGATTAATTTCTCATGTTCTTCTGCTGTAATTTG 540
481 TACATTCGTAGTGTAGTGGAGATTAATTTCTCATGTTCTTCTGCTGTAATTTG 540
541 GGTAAATGATTTGAGTGTGATAGAACCAACCAATTAATCTTACGTTATGAGCTGC 600
541 GGTAAATGATTTGAGTGTGATAGAACCAACCAATTAATCTTACGTTATGAGCTGC 600
601 TTTATATAGTAAAGTCAATTTGTTTCTTCAATACGCAAACTGTTTCGAGAAAGTA 660
601 TTTATATAGTAAAGTCAATTTGTTTCTTCAATACGCAAACTGTTTCGAGAAAGTA 660
661 TAGTAGCATGACAGACTAGATGTAAGACCGAATTCAAATGCTTCTTAAATCTTC 720
661 TAGTAGCATGACAGACTAGATGTAAGACCGAATTCAAATGCTTCTTAAATCTTC 720
721 AAGATTAGACTCTTGAGAGATATGCAATAGAGATGAGCAATGCAAGTTTAACCTCA 780
721 AAGATTAGACTCTTGAGAGATATGCAATAGAGATGAGCAATGCAAGTTTAACCTCA 780
781 GCAGTACAACTGATGCTTATGTCACATCTTCTTCAATGTAATGATGTTTGTGA 840
781 GCAGTACAACTGATGCTTATGTCACATCTTCTTCAATGTAATGATGTTTGTGA 840
841 TCTTCGCTAGTCTTGTGACTTTGCGGGTTCATGAGACCTGATGATGATCTCCATAT 900
841 TCTTCGCTAGTCTTGTGACTTTGCGGGTTCATGAGACCTGATGATGATCTCCATAT 900
901 TCTGTGTAATTTTACATCTTACCAACGTTAAAGTGTGATGTTGTAATCTCTCT 960
901 TCTGTGTAATTTTACATCTTACCAACGTTAAAGTGTGATGTTGTAATCTCTCT 960
961 AGTATGCTGAGTATCAGTCTGATGTCGCAAGCTGATGATGATGATGATGATGAT 1020
961 AGTATGCTGAGTATCAGTCTGATGTCGCAAGCTGATGATGATGATGATGATGAT 1020
1021 GTTCTCTTAGGAGACCTTAATCTTGTGATCAGTTCAGTACGAGAGATTTACCAAGA 1080
1021 GTTCTCTTAGGAGACCTTAATCTTGTGATCAGTTCAGTACGAGAGATTTACCAAGA 1080
1081 CAACTATTTATACAGATATATGTCATATATACCAAAATATGCTGAGGACCAAGAT 1140
1081 CAACTATTTATACAGATATATGTCATATATACCAAAATATGCTGAGGACCAAGAT 1140
1141 ACACCAATTTAGTCAATTTGCAAGTAACCTAATGCTGAGTGAATGACATCTTGAG 1200
1141 ACACCAATTTAGTCAATTTGCAAGTAACCTAATGCTGAGTGAATGACATCTTGAG 1200
1201 ATTATTTGAGACTTTCCATGAGGAGGAAATTTGACTGATGATCAATTAATTTTTC 1260
1201 ATTATTTGAGACTTTCCATGAGGAGGAAATTTGACTGATGATCAATTAATTTTTC 1260
1261 TCCGCTTTGGTGTGATGAGTGTGCTTCTGTAAGCAAAAGAAATTAAGCTTTGTTTC 1320
1261 TCCGCTTTGGTGTGATGAGTGTGCTTCTGTAAGCAAAAGAAATTAAGCTTTGTTTC 1320
1321 TCTGCTTTGAGTGTGATGAGTGTGCTTCTGTAAGCAAAAGAAATTAAGCTTTTATG 1380
1321 TCTGCTTTGAGTGTGATGAGTGTGCTTCTGTAAGCAAAAGAAATTAAGCTTTTATG 1380
1381 GAAACTTTGAGTGTGATGAGTGTGCTTCTGTAAGCAAAAGAAATTAAGCTTTTATG 1440
1381 GAAACTTTGAGTGTGATGAGTGTGCTTCTGTAAGCAAAAGAAATTAAGCTTTTATG 1440
1441 AAGTGAGGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1500
1441 AAGTGAGGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1500

1501 ACCTGACATGGCGAGCAACAAAGGATTCATTTCTTTGTAATTTGATGTTGA 1560
1501 ACCTGACATGGCGAGCAACAAAGGATTCATTTCTTTGTAATTTGATGTTGA 1560
1561 ACACACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
1561 ACACACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
1621 CATGCTCTTCTTAAAGATTAATGATGATGATGATGATGATGATGATGATGATGAT 1680
1621 CATGCTCTTCTTAAAGATTAATGATGATGATGATGATGATGATGATGATGATGAT 1680
1681 CATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
1681 CATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
1741 AGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
1741 AGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
1801 TTGATTTACTTCTGAGATGACCAACAGATGATGATGATGATGATGATGATGATGAT 1860
1801 TTGATTTACTTCTGAGATGACCAACAGATGATGATGATGATGATGATGATGATGAT 1860
1861 TTGGGAAGTCTATGAAACCTGCTCATGAGGATTAACAGCAATTTATGCGCTAACAT 1920
1861 TTGGGAAGTCTATGAAACCTGCTCATGAGGATTAACAGCAATTTATGCGCTAACAT 1920
1921 GATATATGATGAAACAGTCAATTTGATGATGATGATGATGATGATGATGATGATGAT 1980
1921 GATATATGATGAAACAGTCAATTTGATGATGATGATGATGATGATGATGATGATGAT 1980
1981 AGTGAATTAAGAAAGAGACAGATGAGAGATGATGATGATGATGATGATGATGATGAT 2040
1981 AGTGAATTAAGAAAGAGACAGATGAGAGATGATGATGATGATGATGATGATGATGAT 2040
2041 TTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2100
2041 TTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2100
2101 GAGATCCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160
2101 GAGATCCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160

RESULT 2
AAC38032
ID AAC38032 standard; DNA; 1308 bp.
XX AAC38032;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 19544.
XX
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX
XX 05-MAR-1999; 99US-0123180.
XX
XX 09-MAR-1999; 99US-0123348.
XX
XX 23-MAR-1999; 99US-0125788.
XX
XX 25-MAR-1999; 99US-0126264.
XX
XX 29-MAR-1999; 99US-0126785.

Hybridisation assay; genetic mapping; gene expression control;
protein identification; signal transduction pathway;
metabolic pathway; promoter; termination sequence; ss.
Arabidopsis thaliana.

PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130444.
PR 23-APR-1999; 990S-0130510.
PR 23-APR-1999; 990S-0130891.
PR 28-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 30-APR-1999; 990S-0132407.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132486.
PR 06-MAY-1999; 990S-0132487.
PR 07-MAY-1999; 990S-0132863.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0144221.
PR 14-MAY-1999; 990S-0143710.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 17-JUN-1999; 990S-0139453.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140353.
PR 24-JUN-1999; 990S-0140354.
PR 28-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140831.
PR 30-JUN-1999; 990S-014187.
PR 01-JUL-1999; 990S-014187.
PR 01-JUL-1999; 990S-0141842.
PR 02-JUL-1999; 990S-0142154.
PR 06-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.

PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144884.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145224.
PR 23-JUL-1999; 990S-0145276.
PR 26-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 03-AUG-1999; 990S-0146389.
PR 04-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 05-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 06-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 09-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 10-AUG-1999; 990S-0147935.
PR 11-AUG-1999; 990S-0148171.
PR 12-AUG-1999; 990S-0148319.
PR 13-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 16-AUG-1999; 990S-0148684.
PR 17-AUG-1999; 990S-0148368.
PR 18-AUG-1999; 990S-0149175.
PR 20-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0150566.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 01-SEP-1999; 990S-0151303.
PR 01-SEP-1999; 990S-0151308.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155139.
PR 23-SEP-1999; 990S-0155486.
PR 24-SEP-1999; 990S-0155659.
PR 28-SEP-1999; 990S-0156458.
PR 29-SEP-1999; 990S-0156596.
PR 04-OCT-1999; 990S-0157117.
PR 05-OCT-1999; 990S-0157753.
PR 06-OCT-1999; 990S-0157865.
PR 07-OCT-1999; 990S-0158029.
PR 08-OCT-1999; 990S-0158232.
PR 12-OCT-1999; 990S-0158369.

PR	13-OCT-1999;	9905-0159293.
PR	13-OCT-1999;	9905-0159294.
PR	13-OCT-1999;	9905-0159295.
PR	14-OCT-1999;	9905-0159329.
PR	14-OCT-1999;	9905-0159330.
PR	14-OCT-1999;	9905-0159331.
PR	14-OCT-1999;	9905-0159637.
PR	14-OCT-1999;	9905-0159638.
PR	18-OCT-1999;	9905-0159584.
PR	21-OCT-1999;	9905-0160741.
PR	21-OCT-1999;	9905-0160767.
PR	21-OCT-1999;	9905-0160768.
PR	21-OCT-1999;	9905-0160770.
PR	21-OCT-1999;	9905-0160814.
PR	21-OCT-1999;	9905-0160815.
PR	22-OCT-1999;	9905-0160980.
PR	22-OCT-1999;	9905-0160981.
PR	22-OCT-1999;	9905-0160989.
PR	25-OCT-1999;	9905-0161404.
PR	25-OCT-1999;	9905-0161405.
PR	25-OCT-1999;	9905-0161406.
PR	26-OCT-1999;	9905-0161359.
PR	26-OCT-1999;	9905-0161360.
PR	26-OCT-1999;	9905-0161361.
PR	28-OCT-1999;	9905-0161920.
PR	28-OCT-1999;	9905-0161982.
PR	28-OCT-1999;	9905-0161993.
PR	29-OCT-1999;	9905-0162142.

Query Match	24.38;	Score 525.2;	DB 21;	Length 1308;
Best Local Similarity	65.28;	Pred. No. 1.4e-128;		
Matches 1293; Conservative	0;	Mismatches 3;	Indels 686;	Gaps 8

QY	180	ATGAAGAAGATTGAAGAGTGTCCAGGAAGAGACCTTTGG-TTTCAGTTTGTGACT	238
Db	1	ATGAAGAAGATTGAAGAGTGTCCAGGAAGAGACCTTTGGTTTTCAGTTTGTGACT	60
QY	239	CTGTGTGAATGGCTTTTCATGTGAGTTGGGTTTCAATTGGGTCAATGCTTCAAGCA	298
Db	61	CTGTGTGAATGGCTTTTCATGTGAGTTGGGTTTCAATTGGGTCAATGCTTCAAGCA	120
QY	299	CCAAACCCACCTGTTCATATCTGTGCAAGCCCTAATAAGATCAATTACGTTGGAAGG	358
Db	121	CCAAACCCACCTGTTCATATCTGTGCAAGCCCTAATAAGATCAATTACGTTGGAAGG	180
QY	359	AGAAAGAAAGATTTCTACTAGAGTCTCTGCTGTGTGATPAAGAAATTGCAGAGTCT	418
Db	181	AGAAATTAATAATTTCTACTAGAGTCTCTGCTGTGTGATPAAGAAATTGCAGAGTCT	240
QY	419	GCTTAGAAGAAACAGGAAAGATGCTTGTGAACCTTCAATCTCCAGGTATATGCA	478
Db	241	GCTTAGAAGAAACAGGAAAGATGCTTGTGAACCTTCAATCTCCAGGTATATGCA	289
QY	479	ATTACATTTGCTGTAGTGTGAGGAGATTAATTTTCATTTGTTCTGTGCTGAATTT	538
Db	290	-----	289
QY	539	TGGGTAATTTGATTTGAGTGTCTCATTTAGGAACCAACAATTACTTTACTGTTATAGACT	598
Db	290	-----	289
QY	599	GCTTATATTAAGTAAAGATTCAGATTTTGTTTTTCATTCACGAAACGTGTTCCAGAAAG	658
Db	290	-----AG 291	11
QY	659	TATAGTAGCATGACAGAGCTAGATGCTTAAGACCGAACTTCAAAATCTTGTTTAAATCT	718
Db	292	TATAGTAGCATGACAGAGCTAGATGCTTAAGACCGAACTTCAAAATCTTGTTTAAATCT	351
QY	719	TCAAAAGATTACACTCTTGAGAGATGTGCAATAGAGAAATGAGACATGACAGGTTTAACCT	778
Db	352	TCAAAAGATTACACTCTTGAGAGATGTGCAATAGAGAAATGAGACATGACAGGTTTAACCT	400

QY	779	CAGCAGTACAAACGATGCTTTAGTCCCATTTCTTACTTCAATTGATTTGTTGG	838
Db	401	-----	400
QY	839	TATCTGCGTTAAGCTCTTGACTTTGGGGGTTTCATGAGCGCTGAGTATGATCTCCAT	898
Db	401	-----AGTCTTTGACTTTGGGGGTTTCATGAGCGCTGAGTATGATCTCCAT	449
QY	899	ATTCTGTCCTAACTTTTTCACATCTACCAACGTTAACATAGTTGATTTGTAATTCCT	958
Db	450	ATTCTGTCCTAACTTTTTCACATCTACCAACGTTAACATAGTTGATTTGTAATTCCT	497
QY	959	CTAGTTATGCTGAGAGTATACAGTCTGTATTGTCCAAACCTGATGTTCAATATTTTACTGT	1018
Db	498	-----	497
QY	1019	ATGTTCTTCTTTAGGAGCTTAATTCCTTTGCATGATGATGACAGAGATTAACCA	1078
Db	498	-----GGACCTTAACTCTTTGATCAGTGAAGCTGACAGAGCGATTAACCA	543
QY	1079	GACAGATATTAACAGATATATGCTCATATATCACAATATGCTGAGGTGACCACAGA	1138
Db	544	GACAGATATTAACAGATATATGCTCATATATCACAATATGCTGAGGTGACCACAGA	589
QY	1139	ATACACCAAAATTAATCTCAATTCAGATAAACCTAATGCTGAGGTGATAATGATGATCTTG	1198
Db	590	-----	589
QY	1139	AGATTATTTGCAGACTTTCCTCATGAGGAGGAATTTGACTGTAATCCATAAGTTT	1258
Db	590	-----AGACTTTCCATGAGGAGGAGGAATTTGACTGTAATCCATAAGTTT	637
QY	1259	TCTGCGCTTTGGTGATGTGACTATGAGTTTGTCTAGCAAAAGAAAACATTAAGCTTTGT	1318
Db	638	TCTGCGCTTTGGTGATGTGACTATGAGTTTGTCTAGCAAAAGAAAACATTAAGCTTTGT	697
QY	1319	TCTCTGCGCTTTCTAGACTATCATCAAGTATATCTCAGCGCCAAAAGCTAAGTTTAT	1378
Db	698	TCTCTGCGCTTTCTAGACTATAT-----	720
QY	1379	TGGAATCTTGACTGAGAACTATCATCTTTCTCCTCAGGCGATGGCTGAGATGACAT	1438
Db	721	-----CAGCGATGGCTGAGATGACAT	743
QY	1439	CCAAGTGAGAGAGATGAGAACCTCATCTGTGAGAGCCAAATTTGTGAAGACAAACAA	1498
Db	744	CCAAGTGAGAGAGAGATGAGAACCTCATCTGTGAGAGCCAAATTTGTGAAGACAAACAA	803
QY	1499	GTACTGTACATGGCGAGACAAAAGGATTTTCATTTCTTTGTGTAATTTGCATGTTT	1558
Db	804	GTACTGTACATGGCGAGCAAAA-----	826
QY	1559	GAAACGACACTGTATCTGTATTTGTACAAATGATATGATTTGGTGTTTGAGAGATCCTG	1618
Db	827	-----AGGATCCTG	835
QY	1619	GACATGCTCTTTAAAAAGATTAGTAGTGAAAGCAAAAGGTATAAAAAGATTGAT	1678
Db	836	GACATGCTCTTTAAAAAGATTAGTAGTGAAAGCAAAAGCA-----	877
QY	1679	CCCATTAAGTGTCCCATTTATTAATTAAGCTGTGAAGATGTTGAAAATGATTTGAAACAAA	1738
Db	878	-----	877
QY	1739	TCAGAGCTGCTAAGAGGATTTCCGTGTCAATAGGGGTGATAGTTAGGACAAAAACAT	1798
Db	878	--AGAGAGCTGCTAAGAGGATTTCCGTGTCAATAGGGGTGATAGTTAGGACAAAAACAT	935
QY	1799	CATTGATTACTTTCCAGAGTACCAAAAGAAAGATGGAACCTTAAGCGCTAAACAGAGAT	1858
Db	936	CATTGATTACTTTCCAGAGTACCAAAAGAAAGATGGAACCTTAAGCGCTAAACAGAGAT	995
QY	1859	CATTGGAGATCATATGAACCTGTCATCGGATTTTAAACAGAGCAATTTATCGGCTACA	1918

Db	996	CATTGGGAGACGCAATGAAACTCCGACGGATTTTACAGSACATTTATTTGGTAA	1055
QY	1919	ATGATATATGTGAAACAAGTCAGATTTTCAAGTCACTCAACAAGAGCCGTGAACCTTAG	1978
Db	1056	ATGATATATGTGAAACAAGTCAGATTTTCAAGTCACTCAACAAGAGCCGTGAACCTTAG	1115
QY	1979	GAATAGAGATTAACAAAAGCAGCATGAGAGCTCTCAGGTCTATCTGCAATTTCAAGAT	2038
Db	1116	GAATAGAGATTAACAAAAGCAGCATGAGAGCTCTCAGGTCTATCTGCAATTTCAAGAT	1175
QY	2039	GATTTGTTGAGTTACCATGCATCTAGTTTATCAAGGTAGCTCTCAAGCCCTTATCAAA	2098
Db	1176	GATTTGTTGAGTTACCATGCATCTAGTTTATCAAGGTAGCTCTCAAGCCCTTATCAAA	1235
QY	2099	ATGAGATCCCGAGTATGATATATTTAATGAAATGTCTGGTCTCTACCAATCAAA	2158
Db	1236	ATGAGATCCCGAGTATGATATATTTAATGAAATGTCTGGTCTCTACCAATCAAA	1295
QY	2159	CA 2160	
Db	1296	CA 1297	

KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 OS Homo sapiens.
 XX
 XX WO200200928-A2.
 XX
 XX 03-JAN-2002.
 XX
 XX 02-JUL-2001; 2001WO-EP07537.
 XX
 XX 30-JUN-2000; 2000DE-1032529.
 XX 01-SEP-2000; 2000DE-1043826.
 XX
 XX (EPIC-) EPIGENOMICS AG.
 XX
 XX Olek A, Piepenbrock C, Berlin K;
 XX WPI; 2002-130909/17.
 XX
 XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 XX
 XX Claim 1; SEQ ID NO 1645; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 CC
 XX
 XX Sequence 12142 BP; 3009 A; 174 C; 3311 G; 5648 T; 0 other;
 SQ
 Query Match 2.3%; Score 49; DB 24; Length 12142;
 Best Local Similarity 47.3%; Pred. No. 0.032;
 Matches 181; Conservative 0; Mismatches 200; Indels 2; Gaps 1;
 QY 505 GATATATTTCTCATTTGTTCTCTGCTGATTTGGGTAATTTGATTGATTGCTGATT 564
 DB 10246 GTTGTGTTTGTGATTTGTTGTTGTTAGATGTTGTAATAATTTTGGGTTTGT 10305
 QY 565 AGGACCAACCAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 624
 DB 10306 TTTGTGTAATTAATTTGATTGATTGATTGATTGATTGATTGATTGATTGATT 10365
 QY 625 TGTGTTTCTAATCAGAACTGTTTCAGAAAGTATAGCATGACAGAGCTAGATTT 684
 DB 10366 ATTTTATTAAGTTTAAGTTAAGTAGAAGAAATTTGAGAGTTTATAGAGAGAA 10425
 QY 685 TAAGACCGAATCAATGCTGCTTCTTAATCTTCAAGTTAGCATGCTGAGAGATTT 744
 DB 10426 AAGGTGAAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 10483
 QY 745 GGCATATGAGATGAGCAATGAGCTTAATCACTGACAGTACAACTGATTCGTTAGT 804
 DB 10484 TGTGTTGTTGGGGTTTGTGTTTGTGTTTACGATTAATTAATTAATTAATTAATTA 10543
 QY 805 CCCATTCCTTACTTCAATTTGATTGTTGTTGTTATCTTGGCTTACTGCTTTG 864
 DB 10544 CGTTTATTTTGTGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTTT 10603
 QY 865 CGGTTTCATGAGACCTGAGTAT 887
 DB 10604 TTTCGTTTATGGAATTTTAATATG 10626

RESULT 5
 ID ABR40057 standard; DNA; 7441 BP.
 XX
 XX ABR40057;
 AC
 XX
 XX 21-MAY-2002 (first entry)
 DE
 XX Human chemically pretreated gene sequence #70 strand 1.
 DE
 XX Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
 KW cytosinatic; ALDH6; CYP11A; CYP11B; CYP3A3; DPD; EPHX2; OCIN; TXNRD1;
 KW UG8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
 XX
 XX Homo sapiens.
 OS
 XX WO200202806-A2.
 XX
 XX 10-JAN-2002.
 XX
 XX 29-JUN-2001; 2001WO-EP07470.
 XX
 XX 30-JUN-2000; 2000DE-1032529.
 XX 01-SEP-2000; 2000DE-1043826.
 XX
 XX (EPIC-) EPIGENOMICS AG.
 XX
 XX Olek A, Piepenbrock C, Berlin K;
 XX WPI; 2002-154757/20.
 XX
 XX New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
 PT useful for detecting cytosine methylation state of genes associated
 PT with pharmacogenomics and for therapy of diseases e.g. cancer -
 XX
 XX Claim 1; SEQ ID NO 139; 24pp; English.
 XX
 CC The invention relates to a nucleic acid comprising a sequence at
 CC least 18 bases in length of a segment of the chemically pretreated DNA
 CC of genes associated with pharmacogenomics according to one of the
 CC sequences of the genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B
 CC (NM_000497), CYP3A3 (NM_000776 and NM_017460), DPD (NM_000110), EPHX2
 CC (NM_001979), OCIN (NM_002538), TXNRD1 (NM_003330), UG8 (NM_003360),
 CC MRP (NM_004996, NM_019900, NM_019901, NM_019902, NM_019862, NM_019898,
 CC NM_019899) and their complementary sequences, or a sequence (SI) chosen
 CC from 87 sequences and their complements. The chemical pretreatment
 CC is bisulphite treatment to convert cytosines (but not methyl-cytosines)
 CC into uracils. Also included are an oligomer (II) in particular an
 CC oligonucleotide or a peptide nucleic acid (PNA)-oligomer, comprising in
 CC each case at least one base sequence having a length of 9 nucleotides
 CC which hybridises to or is identical to a chemically pretreated DNA of
 CC genes associated with pharmacogenomics and their complements, arranged in
 CC an array for analysing diseases associated with the methylation state
 CC (CpG) and/or detecting SNPs (single nucleotide polymorphisms)
 CC of the 87 sequences. The oligomers may also be used as PCR primers.
 CC The set of 87 nucleic acids and their complements is useful for diagnosis
 CC and therapy of solid tumours and cancer. The present sequence
 CC represents one the 87 DNA sequences or its complement.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 7441 BP; 2098 A; 66 C; 1948 G; 3329 T; 0 other;
 SQ
 Query Match 2.2%; Score 46.8; DB 24; Length 7441;
 Best Local Similarity 42.6%; Pred. No. 0.098;
 Matches 243; Conservative 0; Mismatches 327; Indels 0; Gaps 0;
 QY 91 TGTGTTGCCAATGCGTTTCTCAGCAAGACTGTAATTTATCTCTTTATAGAT 150
 DB 4866 TGTATTTATTTAGATGTTGTAATTAATTAATTAATTAATTAATTAATTAATTTTA 4925

QY 151 AAGAAATCTTGGCTTTTTCAGTTTTCAGTATGAGAAAGAAATGAGAGAGTCCGAGCA 210
 DB 4926 ATTGATTCATTTTATTTTATTTAGATATAGATGAGAAATGAGAAATATATATTTGG 4985
 QY 211 AGGAGACCTTGGTTTCACTTTTGTAGTCTTGTATGAGTCTTATGCAATGAGTTGG 270
 DB 4986 AAAAGAAATGATTTGTTAAAGATAGAGATGTTTGAAGAGACCTTATGTTTGTAT 5045
 QY 271 GTTTCAATGGGTCATGCTTCAAGGCAACCAACCCTGTTCTTAATCTCGAAGCCC 330
 DB 5046 ATAGTTTGTGTTTATTTTGTAGTTTGGAGTTTGGAAAGGTTAGTGTGTTATAGA 5105
 QY 331 TAATAGATCAATTTTCACTTGGAGAGAGAAAGAAAGATTTCTTACTTGAAGTCTGCG 390
 DB 5106 ATATAGTATGAGATAAATATGAGATATTTTGGAAAGATTTGTTAGAGGGTTATGA 5165
 QY 391 TGTGCGATTAAGGATTCGACAGACTGCTTTTGAAGAAACCAAGAAAGATCGTCT 450
 DB 5166 GGTGATTTGTGATATTTTGGTTTATTTTGTATTTTGTATTTAGTAAATATTTGTTAT 5225
 QY 451 TGAACCTTCACATCTCCAGTATATGCAATTTTCAGTCTAGTCTGGAGAGATTAT 510
 DB 5226 TAAATTTAGACTTTATTTATTTTATTTTAAATGATTTTGAAGATTTTGTGAA 5285
 QY 511 ATTTCTCATTTTCTTCTGCTGTAATTTTGGTAAATTTGATTTGATTCATAGGAAC 570
 DB 5286 AATTTTCGATGATGATGATATATGTTTAAATATTTTGGTTATTTTATTTTATA 5345
 QY 571 CAACAATATCTTACGTTATAGACTGCTTATATAGTAAAGTTCAAGTTTGTGTTT 630
 DB 5346 ATTTTAAAGATTTGTTGGAAATTTTATGATTAAGTAAAGTTTGTAAAGATGTAGT 5405
 QY 631 TCTAATCAGCAAACTGTTTCAAGAAAGTA 660
 DB 5406 TTTGAAATAGCAAAATTTATGATGAATATATA 5435

RESULT 6
 AAS63344
 ID AAS63344 standard; DNA; 7479 BP.
 AC AAS63344;
 XX
 DT 29-JAN-2002 (first entry)
 XX
 DE Chemically pretreated metabolism associated gene #39.
 XX
 KW Human; cytostatic; anti-tumour; metabolism; metabolic disease; liver;
 KW solid tumour; cancer; cytosine methylation; epigenetic; eye; kidney;
 KW single nucleotide polymorphism detection; SNP; stool; urine; lung;
 KW cerebral-spinal fluid; intestine; brain; heart; prostate; breast;
 KW DSP2; EBHX; QDPR; SCSH; SHMT2; SLC7A2; SLC7A4; TYMS; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200176451-A2.
 PD 18-OCT-2001.
 PF 06-APR-2001: 2001MO-EP04016.
 PR 06-APR-2000: 2000DE-1019058.
 PR 07-APR-2000: 2000DE-1019173.
 PR 30-JUN-2000: 2000DE-1032529.
 PR 01-SEP-2000: 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-010834/01.
 XX

PT New nucleic acid, useful for diagnosis and therapy of metabolic
 PT disease, solid tumour and cancers, comprises segment of chemically
 PT modified genomic sequences of genes associated with metabolism
 XX
 PS Claim 1; Page 107-109; 143pp; English.

CC The invention relates to a nucleic acid (I) comprising a sequence at
 CC least 18 bases of a segment of the chemically pretreated DNA of genes
 CC associated with metabolism such as DSP2 (NM.004418), EPHX2 (NM.001979),
 CC QDPR (NM.000320), SCSH (NM.000199), SHMT2 (NM.005412), SLC7A2
 CC (NM.003046), SLC7A4 (NM.004173) and TYMS (NM.001071) (all
 CC undefined). (I) are useful for diagnosis and therapy of metabolic
 CC disease, solid tumours and cancers; as primer oligonucleotides for the
 CC amplification of DNA sequences, for detecting the cytosine methylation
 CC state and/or single nucleotide polymorphisms (SNPs) in a chemically
 CC treated DNA of genes associated with metabolism. An array of (I) is
 CC useful for ascertaining genetic and/or epigenetic parameters for the
 CC diagnosis and/or therapy of existing diseases or the predisposition to
 CC specific diseases by analysing cytosine methylations. The method involves
 CC chemically treating genomic DNA sample by a solution of bisulphite,
 CC hydrogen sulphite or disulphite such that cytosine bases which are
 CC unmethylated at the 5th-position are converted to uracil or another base
 CC which is dissimilar to cytosine in terms of hybridisation behaviour and
 CC amplifying fragments of the chemically pretreated genomic DNA. The
 CC genomic DNA is from cells or cellular components which contain DNA,
 CC sources of DNA comprising, for e.g. cell lines, biopsies, blood, sputum,
 CC stool, urine, cerebral-spinal fluid, tissue embedded in paraffin such as
 CC tissue from eye, intestine, kidney, brain, heart, prostate, lung, breast
 CC or liver, histologic object slides and their combinations. Genetic
 CC parameters are mutations, in particular insertions, deletions, point
 CC mutations, inversions and polymorphisms of genes associated with
 CC metabolism and sequences further required for their regulation.
 CC Epigenetic parameters are in particular cytosine methylations and
 CC further chemical modifications of DNA bases of genes associated with
 CC metabolism. Further epigenetic parameters include for e.g. the
 CC acetylation of histones which correlates with DNA methylation.
 CC AAS63306-AAS63373 represent chemically pretreated metabolism associated
 CC genes, and related primers of the invention.
 CC
 XX

SO Sequence 7479 BP; 2111 A; 68 C; 1962 G; 3338 T; 0 other;

Query Match 2.2%; Score 46.8; DB 24; Length 7479;

Best Local Similarity 42.6%; Pred. NO. 0.098;

Matches 243; Conservative 0; Mismatches 327; Indels 0; Gaps 0;

QY 91 TGTGCTGCCAATTCGCTTGTCTCAGTGAACCTGCAAAATTTATCTTTATAGAT 150
 DB 4904 TGTATTTATTTAGATTTGTAGTATATGATTTATTTTAAATTTTATAGTATATTTTA 4963
 QY 151 AAGAAATCTTGGCTTTTTCAGTTTTCAGTATGAGAAAGAAATGAGAGTCCGAGCA 210
 DB 4964 ATTGATTCATTTTATTTTATGATATAGTATGAGAAATGAGAAATATATATTTGG 5023
 QY 211 AGGAGACCTTGGTTTCACTTTTGTAGTCTTGTATGAGTCTTATGCAATGAGTTGG 270
 DB 5024 AAAAGAAATGATTTGTTAAAGATAGAGTCTTTTGAATGACGTTACTTTTGTGAT 5083
 QY 271 GTTTCAATGGGTCATGCTTCAAGGCAACCAACCCTGTTCTATCTTCAAGCCC 330
 DB 5084 ATAGTTTGTGTTATTTTATTTTGTGGAAGTTGTGGAAGGTTGTGTTTATAGA 5143
 QY 331 TAATAGATCAATTTTCACTTGGAGAGAGAAAGAAATTTCTTACTTGAAGTCTGCG 390
 DB 5144 ATATAGTATTTGATTAATATGAAATATTTTGGAAAGACTTTGTTAGAGGGTTATGA 5203
 QY 391 TGTGCGATTAAGGATTCGACAGCTGCTTTTGAAGAAACCAAGAAAGATCGTCT 450
 DB 5204 GGTGATTTGTGATTTTGGTTTATTTTGTAAAGTATGAAATATTTGTTAT 5263
 QY 451 TGAACCTTCACATCTCCAGTATATGCAATTTTCTTACTGATGAGGAGATTAT 510
 DB 5264 TAAATTTAGAGTTTATTTTATTTTAAATGATTTTGAAGATTTTGTGAA 5323

[illegible]

OY 438 AAGAGATCGCTTGAACCTTCACATCCAGGTAATGCAATTAATTCGTAGTGA 497
 DB 9416 AAGAGCTTTTGTTTAAATGTTAATATGTTATATTTGGTAAAGATTTATGATGA 9475
 OY 498 GT--GGGAGGATTAATATTTCATTTGTTTCCTGCTGTCGAATTTTGGTAAATGATTGA 555
 DB 9476 TTATATGTTTTTTTAAATATTTTGTGTTTTTGAAGTTTTTTTATAAAGCTGATATT 9535
 OY 556 GTTGTCATATAGCAACCAACAAATTAATCTTACTGTATAGACTGCTTATATAGTAAAG 615
 DB 9536 TATGCTAATTAATTAATTAAGTAATTAATTAATATTTGTAATATTTGTAATTTTGATT 9595
 OY 616 TTCAGATTTTGTCTTTCTTCTATACAGAAACTGTTTCAGGAAAGATATAGTACAT 669
 DB 9596 TAGTAATTTTATTTTATTTTACGATTAATTTAATTTAAGCAAAATTAATAGATGTAT 9649
 RESULT 8
 ABL70318
 ID ABL70318 standard; DNA; 15224 BP.
 XX ABL70318;
 AC
 DT 01-JUL-2002 (first entry)
 DE Chemically treated cell signalling DNA sequence complementary to#104.
 XX Cell signalling; cytosine methylation; cell signalling disease;
 KM cancer; tumour; cytostatic; ds.
 XX unidentified.
 OS
 PN WO200202807-A2.
 PD 10-JAN-2002.
 XX 29-JUN-2001; 2001WO-EP07471.
 PF 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX (EPIC-) EPIGENOMICS AG.
 PA Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2002-154758/20.
 DR
 XX Nucleic acid, useful for diagnosis and therapy of diseases associated
 PT with cell signalling e.g. cancer, comprises chemically modified genomic
 PI sequences of genes associated with cell signalling -
 PS
 XX Claim 1; SEQ ID NO 208; 24bp+sequence listing; English.
 XX
 CC The invention relates to a nucleic acid comprising a sequence of at least
 CC 18 bases of a segment of chemically pretreated DNA of genes associated
 CC with cell signalling. The activity of the modified sequences of the
 CC invention may be described as cytostatic. The object of the invention is
 CC to provide the chemically modified DNA of genes associated with cell
 CC signalling, as well as oligonucleotides and/or PNA-oligoners for
 CC detecting cytosine methylations, as well as a method which is
 CC particularly suitable for the diagnosis and/or therapy of genetic and
 CC epigenetic parameters of genes associated with cell signalling. The
 CC chemically modified DNA provided by the invention is useful for diagnosis
 CC and therapy of diseases such as solid tumours and cancer. The sequences
 CC given in records ABL70111-ABL70626 represent chemically pre-treated
 CC genomic DNA's of genes associated with cell signalling.
 CC Note: The sequence data for this patent is not represented in the printed
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office.
 XX
 SO Sequence 15224 BP; 3395 A; 417 C; 4284 G; 7128 T; 0 other;

	Query Match	2.18;	Score 46.2;	DB 24;	Length 15224;	
	Best Local Similarity	52.98;	Pred. No. 0.19;			
	Matches 99;	Conservative 0;	Mismatches 88;	Indels 0;	Gaps 0;	
OY	108 TTTTGCTCACTGAAACGTGGAAAAATTTTATCTCTTTTAAGATAAGAATCGTCCTTT	167				
Dd	6170 TTTTTCGGGAAGAGTTACTTTAAGATTTTTTTTTTCGGGATAAATTTTAGATTTTT	6229				
OY	168 TCAGTTTTCAGTTCGAAGAAGAAATTCAGAGAGTCCGAGGAGAACCTTGGTTTC	227				
Dd	6230 TGTGTTTGGAAATTTGGTTGTTTGGGAATTTTGGGGTAGGAATGTTAGTTGGGTTG	6289				
OY	228 AGTTTGTGAGCTCTTGTGTAATGCCTTTATCAATGAGATTGGGTTTCATFTGGCAT	287				
Dd	6290 GGTGTTTATTTTCGAGGTTGCTTGAATTTTTTACTGTGAGTTCGATTTTGTGGGTT	6349				
OY	288 GCCTCAA 294					
Dd	6350 TATTAA 6356					
RESULT 9	AAS61263					
ID	AAS61263 standard; DNA; 15224 BP.					
XX	AAS61263;					
DT	29-JAN-2002 (first entry)					
DE	Human gene regulation-associated gene oligonucleotide #218.					
XX	Human; Gene regulation-associated gene; severe combined immunodeficiency;					
KW	cardiac damage; inflammatory response; Haemophilia; Werner syndrome;					
KW	asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;					
KW	renal disease; Preeclampsia; cardiac allograft vascular disease;					
KW	colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;					
KW	immunostimulant; candid; antinflammatory; coagulant; antiasthmatic;					
KW	nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.					
OS	Homo sapiens.					
XX	WO200177375-A2.					
PN	18-OCT-2001.					
XX	PD					
XX	PF					
XX	06-APR-2001; 2001WO-EP03968.					
PR	06-APR-2000; 2000DE-1019058.					
PR	07-APR-2000; 2000DE-1019173.					
PR	30-JUN-2000; 2000DE-1032529.					
PR	01-SEP-2000; 2000DE-1043826.					
PA	(EPIC-) EPIGNOMICS AG.					
XX	Olek A, Piepenbrock C, Berlin K;					
XX	WT; 2002-017470/02.					
PT	New nucleic acid sequences from chemically modified genes associated					
PT	with gene regulation, useful for analysing cytosine methylations for					
PT	diagnosis and therapy of diseases e.g. severe combined immunodeficiency					
XX	disease					
XX	Claim 1; SEQ ID NO 224; 26pp; English.					
CC	The invention relates to 224 nucleic acid sequences comprising at least					
CC	18 bases of a chemically pretreated gene associated with gene regulation					
CC	selected from 43 known genes (or complementary sequences). The					
CC	chemical pretreatment converts cytosine bases unmethylated at the					
CC	5-position to uracil or another base with hybridisation behaviour					
CC	dissimilar to cytosine, to enable analysis of cytosine methylations.					
CC	The DNA sequences, oligomers (or sets/arrays) and method are					
CC	useful in the diagnosis of diseases (or predisposition to diseases)					

CC	associated with gene regulation and in therapy of such diseases, by
CC	enabling analysis of the cytosine methylation patterns of such genes,
CC	kits are provided. They are especially useful in diagnosis
CC	and therapy of e.g. severe combined immunodeficiency disease, cardiac
CC	disorders, haemophilia, solid tumours and cancer, Werner syndrome,
CC	asthma, HDR syndrome, Saethre-Choizzen syndrome, renal disease,
CC	preclampsia, graft versus-host disease. The present sequence is a
CC	sequence included in the sequence data for this specification and is
CC	associated with the human gene regulation-associated genes.
CC	Note: The sequence data for this patent did not form part
CC	of the printed specification, but was obtained in electronic
CC	format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences
XX	
SQ	Sequence 15224 BP; 3395 A; 417 C; 4284 G; 7128 T; 0 other;
Query Match	2.1%; Score 46.2; DB 24; Length 15224;
Best Local Similarity	52.9%; Pred. No. 0.19;
Matches 99; Conservative	0; Mismatches 88; Indels 0; Gaps 0;
OY	108 TTGGCTCACGAAGACTGTGAAATTTTATCTCTTTATACATAAAGAATCGCTTTT 167
Db	- - - - - - - - - - - - - - -
	6170 TTTTCGAGACGAGTTTAGTTATGAAATTTTTTTTCGGATATATAATTATTAGATTTT 6229
OY	168 TCAGTTTTCAGTATGAGAGAGAAATTTGAAGAGTGTCGCCGAGAGAACCTTGGTTTC 227
Db	6230 TGTTTTTTTGAATTTGGTTGTTTTTGGCAATTTTGGGTAGAAATGTTAGTTGGGTTG 6289
OY	228 AGTTTGAGCTCTGTTGTAATGCCTTTATCATAGAGACTTTGGCTTTCAAFTGGTCAT 287
Db	6290 GGTTTTTTATTCGAGGTTGGTTGATTTTTTTAGTCGAGTTTGATTTTTTTGGGTT 6349
OY	288 GCCTCAA 294
Db	- - - - -
	6350 TATTTAA 6356
RESULT 10	
ID	ABK31359
ID	ABK31359 standard; DNA; 15224 BP.
XX	
AC	ABK31359;
XX	
DT	23-APR-2002 (first entry)
DE	
XX	
Signal transduction associated gene modified complementary DNA #101.	
KW	Human; signal transduction associated gene; cytosine methylation state;
KW	Cpg island; signal transduction associated disease; solid tumour; cancer;
KW	antitumour; cytostatic; mutant; ds.
XX	
Homo sapiens.	
OS	Synthetic.
OS	
WO200200926-A2.	
PN	
PD	03-JAN-2002.
PF	
Ef	29-JUN-2001; 2001WO-EP07472.
XX	
PR	30-JUN-2000; 2000DE-1032529.
PR	01-SEP-2000; 2000DE-1043826.
PA	(EPIg-) EPIGENOMICS AG.
XX	
Pt	Olek A, Piepenbrock C, Berlin K;
DR	WPI; 2002-147896/19.
PT	
signal transductioe for diagnosis and therapy of diseases associated with	
sequences of genes associated with signal transduction	

[illegible]

FT XX PN WO20

TTT 167	PD 12
	PF
TTT 6229	XX 02-2
	XX
QTC 227	PR 04-2
	XX
TTG 6289	PA (YI
	XX
AT 287	PI
	XX
TTT 6349	DR
	WPI

	XX	Nuc
PT	PT	enco
PT	PT	

Class

XX PS XX

CC Lyce
CC ser
CC a no
CC ass
CC are
CC lea
CC in a
CC sub

sub
The
CC
CC
CC

Query 1

Best Lo
Matche

QY

Db

```

exon          /tag= d
              /number= 2
intron        547..879
              /tag= e
              /number= 2
exon          880..1898
              /tag= f
              /number= 3
intron        1899..2656
              /tag= g
              /number= 3
exon          2627..2871
              /tag= h
              /number= 4
intron        2872..2969
              /tag= i
              /number= 4
exon          2970..3066
              /tag= j
              /number= 5
intron        3067..3141
              /tag= k
              /number= 5
exon          3142..3532
              /tag= l
              /number= 6

```

02-JAN-2001; 2001WO-IL00008.
04-JAN-2000; 2000US-0477375.

(YISS) YISSUM RES DEV CO HEBREW

Zamir D, Pleban T, Fridman E;

WPI; 2001-418359/44.
P-PSDB; AAE04416.

Nucleic acid comprising a genome encoding a polypeptide having a function in a particular environment, useful for increasing the growth of a cell in that environment.

Claim 3; Page 93; 99pp; English

The present invention relates to genomic, complementary or compound apoplastic invertase isoforms from *Lycopersicon pennellii* fruit, a serving for secretion into an associated with high brix trait associated with high brix traits are useful for increasing the leaves of plants expressing in a site specific manner, inter substantially increase the The present sequence is Lycopersicon Invertase Isoenzyme.

Very Match	2.18;	Score
1st Local Similarity	52.08;	Pr
atches 102; Conservative	0;	

512 TTTCTCATTTGTTCTTGCTGTC
 ||| |||| |
 2617 TTTTTCATTTATTAACACTTTC

```

OY 572 AACAAATACCTTACTGTTATAGACTGCTTATATAGTAAAGTTGAGATTTGTTTT 631
DB 2557 AAAAAAAAAAACACATTTTATATAGTAAATTTAAAGTACAGTTTCATTTTA 2498
OY 632 CTAATCAGAAACCTTTTCAGAAAGATATAGTACAGACGACTAGATGTAGAGCC 691
DB 2497 TTCTTGTGGACTGCTTAAATTTTAAATAGTACTCAAAATACATTATGAGAGAGAGA 2438
OY 692 GAACCTCAAAATGCTTG 707
DB 2437 AAAGTGAGTATTTTGG 2422

RESULT 12
AAB08701/c
ID AAB08701 standard; DNA; 13431 BP.
XX
AC AAB08701;
XX
DT 04-SEP-2001 (first entry)
XX
XX Lycopersicon pennellii elin5-elin7 invertase isoenzymes genomic DNA.
DE
XX Invertase isoenzyme; apoplast; brix value; monosaccharide; sugar; fruit;
KW green-fruited tomato; elin5; elin7; ds.
XX
XX Lycopersicon pennellii.
OS
XX
FH Key
FH Location/Qualifiers
FT 1..4849
FT /tag= a
FT /label= Lin5_Promoter
FT 4850..8465
FT /tag= b
FT /product= "elin5 invertase isoenzyme"
FT /note= "The CDS is interrupted by 5 introns"
FT 4850..5048
FT /tag= c
FT /number= 1
FT 5049..6332
FT /tag= d
FT /number= 1
FT 6333..6341
FT /tag= e
FT /number= 2
FT /note= "Conserved in plants"
FT 6342..6418
FT /tag= f
FT /number= 2
FT 6419..7440
FT /tag= g
FT /number= 3
FT 7441..7619
FT /tag= h
FT /number= 3
FT /note= "includes an ORF of 30 amino acids (AAB04415)"
FT 7620..7864
FT /tag= i
FT /number= 4
FT 7865..8054
FT /tag= j
FT /number= 4
FT 8055..8154
FT /tag= k
FT /number= 5
FT 8155..8285
FT /tag= l
FT /number= 5
FT 8286..8670
FT /tag= m
FT /number= 6
FT /note= "The above given exons and introns are of
FT Lin5 gene"

```

```

FT misc-feature 8671..9900
FT /tag= n
FT /note= "Unknown region"
FT 9982..13223
FT /tag= o
FT /product= "elin7 invertase isoenzyme"
FT /note= "The CDS is interrupted by 5 introns"
FT 9901..10180
FT /tag= p
FT /number= 1
FT 10181..10436
FT /tag= q
FT /number= 1
FT 10437..10445
FT /tag= r
FT /number= 2
FT 10446..10778
FT /tag= s
FT /number= 2
FT 10779..11797
FT /tag= t
FT /number= 3
FT 11798..12525
FT /tag= u
FT /number= 3
FT 12526..12770
FT /tag= v
FT /number= 4
FT 12771..12868
FT /tag= w
FT /number= 4
FT 12869..12965
FT /tag= x
FT /number= 5
FT 12966..13040
FT /tag= y
FT /number= 5
FT 13041..13431
FT /tag= z
FT /number= 6

PN WO200149826-A1.
XX
XX 12-JUL-2001.
PD
XX
XX 02-JAN-2001; 2001WO-IL00008.
PE
XX
XX 04-JAN-2000; 2000US-0477375.
PR
XX
XX (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
PA
XX
XX Zamlr D, Pleban T, Fridman E;
PI
XX
XX WPI: 2001-418359/44.
DR P-PSDB; AAE04414, AAE04416.
XX
XX Nucleic acid comprising a genomic, complementary or composite sequence
XX encoding a polypeptide having an invertase activity in an apoplastic
XX environment, useful for increasing the monosaccharide content of plants
XX
XX Claim 3; Page 85-88; 99pp; English.
XX
XX The present invention relates to nucleic acids comprising
XX genomic, complementary or composite polynucleotide sequences encoding
XX novel apoplastic invertase isoenzymes associated with high brix value in
XX Lycopersicon pennellii fruit, and an N terminal amino acid sequence
XX serving for secretion into an apoplast. The invention also relates to
XX a novel regulatory element unique to L. pennellii which is also
XX associated with high brix trait. The polynucleotides of the invention
XX are useful for increasing the sugar content in e.g. fruits, roots and
XX leaves of plants expressing them. The regulatory element when integrated,
XX in a site specific manner, into a solanaceae plant genome, can

```


XX WPI: 2002-130909/17.
 DR Nucleic acid comprising fragment of chemically modified gene, useful
 XX for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 PT
 XX
 XX Claim 1: SEQ ID NO 1701; 32pp + Sequence listing; German.
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 CC
 SO Sequence 17721 BP; 4634 A; 328 C; 4081 G; 8678 T; 0 other;
 Query Match 2.1%; Score 45.4; DB 24; Length 17721;
 Best Local Similarity 44.7%; Pred. No. 0.33;
 Matches 223; Conservative 0; Mismatches 271; Indels 5; Gaps 1;
 QY 471 TATATGCAATTACATTCCTAGTGTAGTGGAGGATTATATTCATTGTTCTGCT 530
 DB 11652 TTTTGGATTTTAAATTTTGTATATTTGTTTATAGTGTATATGATGTATATAT 11711
 QY 531 GTGCAATTTGGTAAATTTGATTTGATTCATAGGACCAACAAATTAATCTTACTG 590
 DB 11712 TTTTAAATTTAT-----TATGTTTAAAGGTATGTTTGTAAATTAATAAATAATTAAGGT 11766
 QY 591 TATAGACTGCTATATATAGTAAAGTTCAGATTTGTTTCTATACAGCAACTGTTTC 650
 DB 11767 TTTTTCGTTTGTATATATTAATGATGATTTATTTTAAAGGTAATTTAA 11826
 QY 651 AGGAAAGTATAGTACATGACAGCAGCATGATGTAAGACCGCACTTCAATGCTGCT 710
 DB 11827 AATTAATTTGAAAATTTAGTTTAAGTATGATGGAAGGGGAGTTGATATGTTTAT 11886
 QY 711 TTTAATCTTCAAGATAGTACCTTGAAGGATGCAATAGAAATGAGACAAATCCAG 770
 DB 11887 ATATTTTATTTATTTTGAATATGATAGATATTTTAAATTTGAAAAGAAATA 11946
 QY 771 TTTAATCTGACGCTCAAACTGATGCTTACCTCCATTTCCCTTCAATGATG 830
 DB 11947 TTTAATATTTATTTTGTGAGCTTTGATTTGGAGTTTATTTGATGATAAATTT 12006
 QY 831 ATGTTTGTATCTTCCTAGGCTTGTGACTTTGGCGGTTTCAATGAGCTGATGAT 890
 DB 12007 TGGTTTATTAATTTTATTTGATTAATTAATTTTGTGATTAATTAATTTT 12066
 QY 891 ACACCAATATCTGCTACACTTTTACATCTACACAGCTAAACATAGCTGATTTGAA 950
 DB 12067 AATTAATCTGTTAATTTGAAAATTTTAAATTTTATGATTTAGAAATTTTATTT 12126
 QY 951 GTTATCTCTAGTTAGCT 969
 DB 12127 TATTTGAGCTGTTTGT 12145
 RESULT 15
 ABL34007
 ID ABL34007 standard; DNA; 18434 BP.
 XX
 AC ABL34007;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 1980.
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytosinetic; nootropic;

KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antineuritic; antiarthritic; antidiabetic; antipsoriatic;
 KW antileukemic; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001MO-EP07537.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 XX
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI: 2002-130909/17.
 XX
 XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 PT
 XX
 XX Claim 1: SEQ ID NO 1980; 32pp + Sequence listing; German.
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 CC
 SO Sequence 18434 BP; 5581 A; 233 C; 4257 G; 8363 T; 0 other;
 Query Match 2.1%; Score 45.4; DB 24; Length 18434;
 Best Local Similarity 45.4%; Pred. No. 0.34;
 Matches 163; Conservative 0; Mismatches 196; Indels 0; Gaps 0;
 QY 508 TATATTCGATTTCTCTGCTGATTTTGGTAAATGATTTGCTGCTATGAG 567
 DB 15713 TAGATTAGTTATTTTAAGATTAAGGAGTGTATATTTTAAATTAAGATTTAT 15772
 QY 568 AACCAACAAATTAATCTTACTGTTATAGACTGCTTATATTAAGTAAAGTTCAGATTTTGT 627
 DB 15773 AGAATTTTGAAGTTAAAGTTTAAATTTATTTTAAATTTTAAATTTTAAATTTT 15832
 QY 628 TTTTCTATACAGAACTGTTTACAGAAAGATATGATGATGACAGGACTAGATGTTAA 687
 DB 15833 TTTTATTTTGAAGGCTTTTATGAGACTTATGAGATTAATTAATTAATTAATTTT 15892
 QY 688 GACCAACTCAAAATGCTTCTGCTTAAATCTTCAAGATTAAGTCTGAGAGATTAAGC 747
 DB 15893 GGAAGAAATTTTGAATAAATAATATAGTATAGAGATCGAGATGATTAAGATTA 15952
 QY 748 AATAGAGATGAGCAATGAGGTTTAACTTACAGAGTCAACAACTGATTTGTTAGTCC 807
 DB 15953 ATTAGTAAATTTATTTATTTATTTGATATTAATTAATTAATTTTATTTTGTGCTT 16012
 QY 808 ATTTCCTACTTAAATTTGATTTGTTTGTATTTGCTTGTAGCTTTGACTTTGCG 866
 DB 16013 TAAATTAATTTTAAATTAATTAATTAATTTTATTTTATTTTATTTTATTTTGTAG 16071
 Search completed: June 24, 2003, 12:40:21
 Job time : 342 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 11:46:34 ; Search time 2052 Seconds
(without alignments)
17047.881 Million cell updates/sec

Title: US-09-870-406a-32
Perfect score: 2160
Sequence: 1 gaattccccacgcacacgtg.....tcgtctctactaatcaaca 2160

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 809774376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hnc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hnc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rnd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	396	18.3	396	10	AV814470 AV814470
2	373	17.3	417	9	AT999062 70151696
3	309	14.3	309	17	AL081825 Arabidops
4	111	5.1	111	17	AO011370 F7D23TRC
5	101.6	4.7	404	12	BG544889 E2953 Ch1
6	93.6	4.3	606	13	BM10119 EST557655

7	93	4.3	285	17	BH543454
8	90.6	4.2	449	10	AW234724
9	86.4	4.0	410	10	AW234631
10	81.4	3.8	701	14	BQ999997
11	78.4	3.6	372	10	BE059899
12	77.4	3.6	519	13	BM038189
13	75.8	3.5	516	17	AZ128578
14	73.2	3.4	547	12	BE940987
15	71.2	3.3	645	10	BE611353
16	69.4	3.2	525	13	BM382166
17	63	2.9	734	13	BJ169198
18	62.2	2.9	668	12	BG127982
19	61.4	2.8	547	14	BM887362
20	61.2	2.8	1101	17	CNS00396
21	60.8	2.8	547	13	BI427080
22	55.4	2.6	1101	17	CNS00396
23	54.2	2.5	620	12	BG097819
24	51	2.4	434	13	BM521774
25	51	2.4	624	10	AW756255
26	48.8	2.3	1101	17	CNS008X5
27	48.4	2.2	1074	17	CNS0152R
28	48.2	2.2	1101	17	CNS0106X
29	48.2	2.2	1141	17	CNS07206
30	48	2.2	1101	17	CNS00262
31	47.8	2.2	996	17	CNS00F0H
32	47	2.2	919	17	AZ542812
33	47	2.2	1101	17	CNS0177R
34	46.8	2.2	1101	17	CNS0106X
35	46.6	2.2	821	17	BH453266
36	46.4	2.1	1101	17	CNS0006J
37	45.8	2.1	1101	17	CNS00L8K
38	45.8	2.1	1101	17	CNS00E8T
39	45.8	2.1	1101	17	CNS00FVE
40	45.8	2.1	1101	17	CNS00L72
41	45.4	2.1	906	17	AZ531298
42	45.4	2.1	928	17	CNS00DKY
43	45.4	2.1	964	17	CNS07E8R
44	45.4	2.1	1101	17	CNS016L1
45	45	2.1	1001	17	CNS01400

ALIGNMENTS

RESULT 1
AV814470/c
LOCUS
DEFINITION
AV814470 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-83-B22 3',
mRNA sequence.
ACCESSION
AV814470
VERSION
AV814470.1 GI:19856262
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana
thale cress.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidops.
REFERENCE
1 (bases 1 to 396)
Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
Arahawa,T., Shibata,K., Shingawa,A., Muramatsu,M., Hayashizaki,Y.
and Shinozaki,K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished (2002)
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rct.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI

SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana
REFERENCE Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
AUTHORS 1 (bases 1 to 309)
Salanoubat, M., Choisme, N., Artiguenave, F., Brothier, P., Wincker, P., Samson, D., Saurin, W., Weissenbach, J. and Quetier, F.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 309)
Genoscope.
AUTHORS Direct Submission
TITLE Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
FEATURES
Source Location/Qualifiers
1. 309
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="F3124"
/clone_lib="IGF"
/note="end : SP6"
BASE COUNT 73 a 56 c 75 g 105 t
ORIGIN

Query Match 14.3%; Score 309; DB 17; Length 309;
Best Local Similarity 100.0%; Pred. No. 1.4e-65;
Matches 309; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CCCACGTCACAGCTACTGTCATTCACAGTCGCGAGTGTGGCCCTATAGTTGACACATG 66
1 CCCACGTCACAGCTACTGTCATTCACAGTCGCGAGTGTGGCCCTATAGTTGACACATG 60
DB 1
QY 67 ACTGGAGCAGATGTTGAATTCATTCGCTTCCCAATTCGCTTGTCTCACTGAAACTGT 126
61 ACTGGAGCAGATGTTGAATTCATTCGCTTCCCAATTCGCTTGTCTCACTGAAACTGT 120
DB 61
QY 127 GAAATTTTATCTCTTTATGATAAGAAATCTGCTTTTTCAGTTTTCAGTATGAGA 186
121 GAAATTTTATCTCTTTATGATAAGAAATCTGCTTTTTCAGTATGAGA 180
DB 121
QY 187 AGAATTTGAAGAGTGTCCGAGAGAGACCTTGTGTTGAGTTGTGAGTCTGTGT 246
181 AGAATTTGAAGAGTGTCCGAGAGAGACCTTGTGTTGAGTTGTGAGTCTGTGT 240
DB 181
QY 247 AATGCTTTATCAATGAGATTTGGGTTTCAATTGGGTCATGCTTCAAGGACCAAAACC 306
241 AATGCTTTATCAATGAGATTTGGGTTTCAATTGGGTCATGCTTCAAGGACCAAAACC 300
DB 241
QY 307 ACCGTCTCT 315
301 ACCGTCTCT 309
DB 301

RESULT 4
LOCUS A0011370/c 111 bp DNA linear GSS 29-MAY-1998
DEFINITION F27D23TFC IGF Arabidopsis thaliana genomic clone F27D23, DNA
sequence.
ACCESSION A0011370
VERSION A0011370.1 GI:3166615
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 111)
Rounsley, S.D., Suh, E.J., Mable, C., Golden, K., Shatsman, S., Choi, P., Yu, K., Akintreloye, B., Shen, K., Goonasekaram, S., Millscher, J., Adams, M.D. and Venter, J.C.
AUTHORS A BAC End Sequence Database for Identifying Minimal Overlaps in
TITLE

JOURNAL Arabidopsis Genomic Sequencing. Update 4
COMMENT Unpublished (1998)
Other GSSs: F27D23TFC
Contact: Steve Rounsley
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: rounsley@ligr.org
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 111.
FEATURES
Source Location/Qualifiers
1. 111
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="F27D23"
/clone_lib="IGF"
/sex="hermaphrodite"
/note="Vector: BeLOBACII; Site_1: EcoRI; Site_2: EcoRI;
Produced by Thomas Altmann"
BASE COUNT 29 a 21 c 24 g 37 t
ORIGIN

Query Match 5.1%; Score 111; DB 17; Length 111;
Best Local Similarity 100.0%; Pred. No. 8.8e-17;
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 AAGGACCAAAACCCACCTGTTCTATCTCTGCAAGCCCTAATAGATCAATTCACGTTG 352
111 AAGGACCAAAACCCACCTGTTCTATCTCTGCAAGCCCTAATAGATCAATTCACGTTG 52
DB 111
QY 353 AGAAGGAGAAAGAAAGATTTCTTACTAGAGTCTGCTGTCTCTATPAG 403
51 AGAAGGAGAAAGAAAGATTTCTTACTAGAGTCTGCTGTCTCTATPAG 1
DB 51

RESULT 5
LOCUS BG544889 404 bp mRNA linear EST 01-MAY-2002
DEFINITION E2953 Chinese cabbage etiolated seedling library Brassica rapa
subsp. pekinensis cDNA clone E2953, mRNA sequence.
ACCESSION BG544889
VERSION BG544889.1 GI:20375870
KEYWORDS EST.
SOURCE Brassica rapa subsp. pekinensis.
ORGANISM Brassica rapa subsp. pekinensis
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 404)
Bank, J.D., Lee, S.-Y., Kang, J.S., Kim, C.-H., Kim, C.-Y., Choi, Y.-J., Lee, S.-H., Ryu, S.-H., Lee, S.-Y., Cho, M.-J. and Lim, C.-O.
AUTHORS Expressed Sequence Tags of Chinese Cabbage Etiolated Seedling cDNA
Unpublished (2001)
TITLE JOURNAL
COMMENT
CONTACT: LHM, C.O.
Plant Molecular Biology & Biotechnology Research Centre
Gyeongsang National University
#900 Gazeo-dong, Jinju 660-701, Korea
Tel: 82 55 751 6255
Fax: 82 55 759 9363
Email: collim@hongae.gsnu.ac.kr
Seq primer: 77.
FEATURES
Source Location/Qualifiers
1. 404
/organism="Brassica rapa subsp. pekinensis"
/cultivar="Jangwon"
/db_xref="taxon:51351"
/clone="E2953"
/clone_lib="Chinese cabbage etiolated seedling library"
/tissue_type="Etiolated seedling"

```

/lab host="XL-1 Blue"
/note="Vector: pSPORT 1; Site_1: Sal I; Site_2: Not I"
BASE COUNT      117 a      94 c      89 g      104 t
ORIGIN

Query Match
Best Local Similarity 80.4%; Score 101.6; DB 12; Length 404;
Matches 119; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 626 GTTTTTCATCAGAACTGTTTCAGAAAGATATAGATGACAGAGACTAGAGT 685
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 76 GTGCTTTCAGCTTCACCTCCAGGAAAGATAGTACATGACAGAGACTGATGCC 135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 686 AAGACCGAACTTCAAGCTGCTTTTAAATCTTCAAGATTAAGCTTGGAGATG 745
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 136 AAAACACAACCTTAATCTGCTTCAAACTCCAAAGATTAGACTCTTGGGAGATG 195
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 746 GCATAGAGATGAGACATGAGGTTT 773
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 196 TCATTTGAGACGAGACATGACGTGT 223
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
LOCUS      BM110119      606 bp      mRNA      linear      EST 26-NOV-2001
DEFINITION EST557655 potato roots Solanum tuberosum cDNA clone cPR06114 5' end
ACCESSION  BM110119
VERSION     BM110119.1 GI:17071552
KEYWORDS   EST.
SOURCE     potato.
ORGANISM   Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; eusterids I; Solanales; Solanaceae; Solanum.
REFERENCE  1 (bases 1 to 606)
            Van der Hoeven R., Sun, H., Karanycheva, S.A., Tsai, J., Van Aken, S.,
            Uterback, T., Chlemlung, A., Bougri, O., Buell, C.R., Ronning, C.,
            Tanksley, S. and Baker, B.
            Generation of ESTs from potato roots
            Unpublished (2001)
            Contact: Research Genetics, Libraries Division
            Tel: 1-800-711-6195
            Email: cdna@resgen.com
            For clone info: please contact Research Genetics, Libraries
            Division tel 1-800-711-6195, email cdna@resgen.com
            Seq primer: T3.
FEATURES
    source
        1..606
            location/Qualifiers
                /organism="Solanum tuberosum"
                /cultivar="Kennedec"
                /db_xref="taxon:4113"
                /clone="cPR06114"
                /clone_lib="potato roots"
                /tissue_type="roots"
                /dev_stage="in vitro grown stem cuttings"
                /lab_host="SOLR"
                /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                XhoI; supplier: Cornell University, Tanksley lab;
                sequencing: The Institute for Genomic Research. Roots were
                isolated from in vitro grown stem cuttings on CM medium.
                Roots were isolated two weeks after placing the stem
                cuttings from in vitro grown plants on medium."
BASE COUNT      186 a      96 c      138 g      186 t
ORIGIN

Query Match
Best Local Similarity 66.2%; Score 93.6; DB 13; Length 606;
Matches 135; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 1741 AGAGCTGCTAAGGATTCCTGTTCAATGGGTGATGAGTTAGGACAAAACATTA 1800
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 299 AGAATGTGATGGGAACTTCATTCATGATGTTAGTATCTTGGGACAAAACATTTTC 358
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 1801 TTGATTAATCTTCCAGACTACCAAGAGATGAGTACTGTAAAGCATTAACGAATGTA 1860
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 359 TTGATTAATCTTCCAGACTACCAAGATGAGTACTGTAAAGCATTAACGAATGTA 418
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1861 TTGGAGAGTATATGAATCTGTCATGAGATTTTACAGACAAATTTATGCGCTAACAT 1920
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 419 TCGGAATATCTTCGAAATCGAACCATGAGATCGAGAGAAATTCATGATGATAGAG 478
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1921 GATATATGTGACAGATCAGATTT 1944
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 479 TTAAGATAGTCTGATGTAATATTT 502
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
LOCUS      BH543454      285 bp      DNA      linear      GSS 14-DEC-2001
DEFINITION BOGDJ52TR BOGD Brassica oleracea genomic clone BOGDJ52, DNA
ACCESSION  BH543454
VERSION     BH543454.1 GI:17795235
KEYWORDS   GSS.
SOURCE     Brassica oleracea.
ORGANISM   Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE  1 (bases 1 to 285)
            Town, C.D., Van Aken, S., Uterback, T. and Fraser, C.M.
            Whole genome shotgun sequencing of Brassica oleracea
            Unpublished (2001)
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES
    source
        1..285
            location/Qualifiers
                /organism="Brassica oleracea"
                /strain="TO1000DH3"
                /db_xref="taxon:3712"
                /clone="BOGDJ52"
                /clone_lib="BOGD"
                /note="Vector: pPOS1; Site_1: BstXI; 2-3 kb sheared
                genomic DNA inserted into pPOS1 using BstXI linkers"
BASE COUNT      71 a      50 c      58 g      106 t
ORIGIN

Query Match
Best Local Similarity 87.2%; Score 93; DB 17; Length 285;
Matches 102; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 848 TTAGCTTTGACTTTGGGGTTTCATGGAGCGCTGAGATGATATCCCATTTCTGTC 907
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 83 TTAGGCTTTGACTTTGCTGTTTATGAGGCTGATGATGATACCTATATTTCTGTC 142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 908 TAACCTTTTCACATCTACCAAGCTTAACATAGTTTATGTAAGTTATCTTCTACTT 964
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 143 TAACCTTTTCACATCTACCAAGCTTAACATAGTTTATGTAAGTTATCTTCTACTT 199
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
LOCUS      AM234724      449 bp      mRNA      linear      EST 17-JUL-2000
DEFINITION sfl8e08.y1 Gm-c1028 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
ACCESSION  AM234724
VERSION     AM234724.1 GI:6567097
KEYWORDS   EST.

```



```

/ab_xref="taxon:4236"
/clone="OGG3K04"
/clone_lib="QG_EFGHJ lettuce scirriola"
/lab_host="E.coli"
/note="vector: pBRCDNAStIAB: The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgpdb.ucdavis.edu/
TMG_LIB-QG_EFGHJ lettuce scirriola
TMG_TISSE-flowers pre-fertilized
TMG_SEO-GCTTGACGG"

```

	Query March Best Local Similarity	3.8%; Score 81.4; DB 14'	Length 701;
	Matches 112; Conservative 0; Mismatches 51; Indels 0; Gaps 0;		
QY	1186 ATGACTGATCTTGAGATTATTTGAGACATTTCCCATGGGGAGAGAAATTCAGTGTGAA	1246	
Db	448 ATTCCCTTGATGTGCAATCTAGACTTCGAGCTTTGGCATGGGGAGAGAAATTCAGTGTGAA	507	
QY	1246 TCCATTAATTTTTTTCGCCCTTTGGTATGTGGCATGAGTTTTCGCTACGCAAAAAA	1305	
Db	508 TCGTGAATTTTTTTTTCACCAATGTATATATGGACAAAGTTCTTCAAGCCAAAGAAAC	567	
QY	1306 CATAAGGCTTTGCTCTGCGCTTCTGAGAGACATACAGGTAT	1348	
Db	568 CACAATATATTTGTTTCCGCTTTCAAGATATACACAGGCAT	610	

[illegible]

1 (bases 1 to 372)
Shomaker, R., Kelm, P., Vodka, L., Erpelting, J., Coryell, V., Rhanna
A., Bolla, B., Marra, M., Hillier, L., Kucanas, T., Martin, J., Beck, C.
Wyllie, T., Underwood, K., Stepeco, B., Theising, B., Allen, M., Bowers
Y., Person, B., Swaller, T., Gibbons, M., Pepe, D., Harvey, N., Schurk
R., Ritter, E., Kohn, S., Shih, T., Jackson, Y., Cardenas, M., McCann
R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)

```
/organism="Lactuca sativa"  
/cultivar="L.serriola"
```

Email: estewatson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4463 or contact via email: ccu@resgen.com
Insert length: 105 Std Error: 0.00
Seq primer: -40RP from Gibco
High quality sequence stop: 361.

Location/Qualifiers
1. .372

```

/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1016-13799"
/clone_lib="Gm-c1016"
/tissue_type="mature flowers of field grown plants"
/lab_host="Xl10-Gold"
/notes="Vector: pluescript II XR; Site.1: EcoRI; Site.2: XhoI; This cDNA library was constructed from mRNA isolated from immature flowers of field grown plants. The cDNA library was prepared using the Stratagene pluescript II XR library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pluescript vector. The ligated cDNA fragments were transformed into Xl10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelding."

```

Query Match	Best Local Similarity	3.6% ; Score 78.4 ; DB 10 ; Length 372 ;
Matches 106 ;	Conservative 0 ; Mismatches 46 ; Indels 0 ; Gaps 0 ;	
QY	1741 AGGAGCGCTAAGGATTTCCCTTCATGGGGTGATGAGTTAGGCACAAAACATTC	1800
Db	216 AGGATATGTGAGAAGCTTCTCTTATATGAGTCATGAACTTGGAAAGCAAAATTC	275
QY	1801 TTGATTACTTCCAGATACCAACAGAGATGGAAGCTGTAAAGGATTAACGAATATCA	1860
Db	276 ATGATTATTTCCACAGTACTGTCTGCAAGAGGAAATCTTAATTAATAAAGCAATATTA	335
QY	1861 TTGGGAAGTCATATGAAGATCGTCATNGGAT	1892
Db	336 TTGGGAAGTCCTTTGAAAATCGCCCTTGAGAT	367

FEATURES	SOURCE
LOCUS	BM038189
DEFINITION	519 bp mRNA linear EST 06-NOV-2001
ACCESSION	U005B12
VERSION	Oryza sativa mature leaf library induced by M.grisea Oryza sativa cDNA clone U005B12, mRNA sequence.
KEYWORDS	BM038189
ORGANISM	BM038189.1 GI:16753810
SOURCE	EST.
ORyza sativa.	
ORyza sativa	
Eukaryota:	Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 519)	
Dong,H.T., Li,D.B., Zhuang,X.F., Dai,C.G., Sun,L.X., Pei,Y.X., Wu,H.F., Jiang,Y.X., Yu,F.C., Gao,Q.K. and Lou,Y.C.	
A Gene Expression Screen in Oryza sativa	
Unpublished (2001)	
Contact: Dong HT	
Laboratory of Functional Genetics	
Bio-technology Institute of Zhejiang University	
Kaixuan Road 268#, Hangzhou, Zhejiang, P.R.China	
Tel: 0086-571-86892051	
Fax: 0086-571-86961525	
Email: htdong@zjuem.zju.edu.cn	
Seq primer: M13 forward primer.	
Location/Qualifiers	
1..519	

```

/clone.lib="Oryza sativa mature leaf library induced by
M.grisea"
/tissue.type="leaf"
/dev.stage="Mature stage"
/notes="Vector: pSport2"
BASE COUNT      164 a      95 c      109 g      150 t      1 others
ORIGIN

```

	Query Match	Best Local Similarity	Score 6.8;	Score 7.4;	DB 13;	length 519;
			65.1%;	Pred. No. 2.8e-08;		
	Matches 114;	Conservative 0;	Mismatches 61;	Indels 0;	Gaps 0;	
QY	1741	AGGAGTCGTAAAGGATTCCTGTTCATATGGGGTGGATGAGTTAGGCCAATAAATTC	1800			
Db	181	AGGATCTGGTGATGGAGTTCTGTGGTAAGGAGTAATACTCTTGGAAATAATCATTTTC	240			
QY	1801	TTGATTTACTTCCAGAGTACCAACAGAAAGTGAAGTGTAAAGCATTAACGAAGTATCA	1860			
Db	241	TGGATTTATTTCCCTGGAATACGCCCGCTGATGTGGGTCAAGTAAACAGAAAAGGAGTATGA	300			
QY	1861	TTGGAGATCATATGAAGTCGTCATGGAGATTAAACAGCAATTTATCGGCTA	1915			
Db	301	TGGCAAGTCTTTTGAATACTAGCGCTTGGAGCCCGCCCATGAGAAATTCATTGGCA	355			

FEATURES	SOURCE
LOCUS	AZ128578/c
DEFINITION	OSINBD0099G12r CUGI Rice BAC library (EcorI) Oryza sativa genomic clone OSINBD0099G12r, DNA sequence.
ACCESSION	AZ128578
VERSION	AZ128578.1 GI:8205983
KEYWORDS	GSS.
SOURCE	Oryza sativa.
ORGANISM	Oryza sativa
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
AUTHORS	1 (bases 1 to 516)
TITLE	Wing, R.A. and Dean, R.A.
JOURNAL	A BAC End Sequencing Framework to Sequence the Rice Genome
COMMENT	Unpublished (1998)
	Contact: Wing RA
	Clemson University Genomics Institute
	Clemson University
	100 Jordan Hall, Clemson, SC 29634, USA
	Tel: 864 656 7288
	Fax: 864 656 4293
	Email: twing@clemson.edu
	Seq primer: GGAACAGCTGTGACCATG
	Class: BAC ends
	High quality sequence start: 28
	High quality sequence stop: 487.
	Location/Qualifiers
	1..516

```

/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="OSJNB00099612r"
/clone_lib="CGI Rice BAC Library (EcoRI)"
/risuse_type="Leaf"
/1ad_host="E. coli DH10b"
/note="Vector: pBACindigo; Site_1: EcoRI; Site_2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
nonocryledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and
Earle, 1991). The relatively small genome of rice, three
times larger than that of Arabidopsis, makes it suitable
for genomic studies. In order to facilitate positional

```

3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527

```
Matches 102; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
```

Matches 110; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

[illegible]

1795 CATTCATTGATTACTTCCAGAGTACCAACAGAGATGGAAGTCTAAGCGATTAACGAA 1854

1855 GTATCATTTGGGAAGTCATATGAACTCGTCGATGGGATTTAACAGCA 1901

Appendix 1

DEFINITION	EST420566	MGHG	Medicago truncatula	cdna clone pmGHG-2020, mRNA	EST 03-0C1-2000
------------	-----------	------	---------------------	-----------------------------	-----------------

KEYWORDS: EST, barra, medic

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae: eurosoids I; Fabales; Fabaceae; Papilionoideae; Vicia

AUTHORS
Cote, F., Ojanen-Reuths, T., Hahn, M.G., Vandenbosch, K., Hur, J.,
Beremand, P., Endre, C., Toun, C. D., Portier, C. I., Gaudin, M. J., et al.

With beta glucan elicitor preparation from *Phytophthora sojae*
[Published (2000)]

University of Georgia
220 Riverbend Road, Athens, GA 30602-4712, USA

• Email: hahn@ccrc.uga.edu
University of Georgia name: G268839e TIGR sequence name: MTAB047K

```
source      1. .547
            /organism="Medicago truncatula"
```

```
/clone="pmGHC-2020"  
/clone_lib="MGHC"
```

```
elicitor preparation from Phytophthora sojae"
/lab_host="E. coli strain XLOLR"
```

Db 1 AACGAGTCGATGAACTATGAAGCAAAACCTTCAATGACTATTTCTCATTTACCACGGC 60

[illegible]

1887 TGGGATTTTAAACAGGACCAATTTATCCGCTAA 1915

LOCUS	BE611353	645 bp	mRNA	linear	EST	06-DEC-2001
DEFINITION	sg80a01.y1	Gm-cl048	glycine max	cdna clone	GENOME SYSTEMS	CLONE ID:

ACCESSION	BE611353
VERSION	BE611353.1
	GI:9902385

ORGANISM
glycine illax
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

glycine.
1 (bases 1 to 645)

Wyrie, I., Underwood, R., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk

•
•
FABRIC SOYBEAN ESI PROJECT
UNPUBLISHED (1999)
JOURNAL

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

This clone is available through: ResGen, Invitrogen Corp. 2130 South Memorial Parkway, Rockville, MD 20850

	source
1.	.645

```
/clone_lib="Gm-cl048"
```

```
/note="Vector: pBluescript II Sk+; Site_1: EcoRI; Site_2:
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 14:18:39 ; Search time 3711 Seconds

(without alignments)
16939.386 Million cell updates/sec

Title: US-09-870-406a-32

Perfect score: 2160

Sequence: 1 gaattccccacgtcaacgtg.....tcgtctctactaatcaaca 2160

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl1.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sy.*
12: gb_un.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vl.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rnd.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2160	100.0	85561	8	ATAC009326	AC009326 Arabidops
2	2160	100.0	106688	8	ATAC011436	AC011436 Arabidops
3	406	18.8	1358	8	AY099706	AY099706 Arabidops
4	381	17.6	1337	8	AB045112	AB045112 Arabidops
5	318	14.7	1308	8	AY088874	AY088874 Arabidops
6	278	12.9	1092	8	AY128900	AY128900 Arabidops
7	26	1.2	85561	8	ATAC009326	AC009326 Arabidops
8	23	1.1	2674	10	GP1P4511	D11043 Cavia porce
9	22	1.0	1077	10	S79865	S79865 Pig-A-synth
10	22	1.0	101305	10	AL663025	AL663025 Mouse DNA
11	22	1.0	165720	10	AL732475	AL732475 Mouse DNA
12	22	1.0	170821	10	AL645626	AL645626 Mouse DNA
13	22	1.0	176725	2	AC115692	AC115692 Mus muscu
14	22	1.0	221729	2	AL844855	AL844855 Mus muscu
15	22	1.0	228443	2	AC125049	AC125049 Mus muscu
16	21	1.0	1011	14	AF013368	AF013368 Carey Isl
17	21	1.0	1303	8	AY051037	AY051037 Arabidops
18	21	1.0	1472	8	AF360275	AF360275 Arabidops
19	21	1.0	5589	9	AF523304	AF523304 Homo sapi
20	21	1.0	39634	3	AF045641	AF045641 Caenorhab
21	21	1.0	55411	2	AC100139	AC100139 Mus muscu
22	21	1.0	64120	9	AL445646	AL445646 Human DNA
23	21	1.0	85907	8	ATP28M11	ATP28M11 Arabidops
24	21	1.0	92495	8	ATP5L19	ATP5L19 Arabidops
25	21	1.0	119012	2	AC094304	AC094304 Rattus no
26	21	1.0	119926	2	AC123166	AC123166 Rattus no
27	21	1.0	127397	2	AC122160	AC122160 Medicago
28	21	1.0	127400	2	AC131121	AC131121 Mus muscu
29	21	1.0	141984	8	F9L1	AC007591 Arabidops
30	21	1.0	147811	2	AC121389	AC121389 Rattus no
31	21	1.0	150863	2	AC116216	AC116216 Rattus no
32	21	1.0	152141	2	AC119787	AC119787 Rattus no
33	21	1.0	160867	9	AC025254	AC025254 Homo sapi
34	21	1.0	163749	2	AC111796	AC111796 Rattus no
35	21	1.0	166976	2	AC034274	AC034274 Homo sapi
36	21	1.0	170532	9	AL356421	AL356421 Human DNA
37	21	1.0	172507	9	AC092960	AC092960 Homo sapi
38	21	1.0	174811	2	AC125144	AC125144 Mus muscu
39	21	1.0	177063	2	AC069211	AC069211 Homo sapi
40	21	1.0	178106	2	AC127654	AC127654 Rattus no
41	21	1.0	180280	2	AC094064	AC094064 Rattus no
42	21	1.0	183544	2	AC112631	AC112631 Rattus no
43	21	1.0	183663	2	AC124842	AC124842 Rattus no
44	21	1.0	185945	2	AC120940	AC120940 Rattus no
45	21	1.0	198005	8	ATCHRIV28	AL161516 Arabidops

ALIGNMENTS

RESULT 1
ATAC009326 85561 bp DNA linear PLN 24-JAN-2001
LOCUS Arabidopsis thaliana chromosome III P1 MZB10 genomic sequence,
DEFINITION Complete sequence.
ACCESSION AC009326
VERSION AC009326.8 GI:12408713
KEYWORDS HTG.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
REFERENCE Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidops.
1 (bases 1 to 85561)
AUTHORS Lin, X., Kaul, S., Town, C.D., Bente, M.-I., Creasy, T.H., Haas, B.,

Roaring, C.M., Koo, H., Fujii, C.Y., Uterback, T.R., Barnstead, M.E.,
Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.
Arabidopsis thaliana chromosome III p1 MZB10 genomic sequence
2 (bases 1 to 85561)
lin.x. and Kaul, S.
TITLE
JOURNAL
Submitted (16-AUG-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA, xlinetlgr.org
3 (bases 1 to 85561)
lin.x.
Direct Submission
Submitted (24-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
On Jan 24, 2001 this sequence version replaced g1:12280842.
Address all correspondence to:
Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr.
Rockville, MD 20850, USA
e-mail: xlinetlgr.org
P1 clone MZB10 is from Arabidopsis chromosome III and is near the
molecular marker C1C7A12.
The orientation of the sequence is from SP6 to T7 end of the P1
clone.

COMMENT

Genes were identified by a combination of three methods: Gene
prediction programs including GRAIL (available by anonymous ftp
from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of
Washington), Genscan (Chris Burge,
http://genomic.stanford.edu/~chris/GENSCANV.html), and NcapIntGene
complete sequence against a peptide database and the Arabidopsis
EST database at TIGR (http://www.tigr.org/tldb/at.html).
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity, that are predicted by more
than two gene prediction programs over most of their length are
annotated as 'hypothetical proteins'. Genes encoding tRNAs are
identified by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
identified by RepeatMasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
genomic sequence that are not annotated as genes but have predicted
exons by GRAIL are annotated as misc features.
Location/Qualifiers
1..85561

FEATURES

source

/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="III"
/map="C1C7A12"
/clone="MZB10"
349..702
misc_feature
/note="exon predicted by xgrail, quality excellent"
complement(587..681)
/note="exon predicted by xgrail, quality
excellent.shadowexon"
881..1171
misc_feature
/note="exon predicted by xgrail, quality good"
1508..1606
misc_feature
complement(1516..1602)
/note="exon predicted by xgrail, quality
marginal.shadowexon"
1655..1758
misc_feature
/note="exon predicted by xgrail, quality marginal"
complement(1682..1731)
/note="exon predicted by xgrail, quality
marginal.shadowexon"
complement(1852..1922)
/note="exon predicted by xgrail, quality good_shadowexon"

misc_feature
1854..1933
/note="exon predicted by xgrail, quality marginal"
<2185..>3281
/gene="MZB10.1"
/note="similar to MITOCHONDRIAL INNER MEMBRANE PROTEASE
SUBUNIT 2 GB:P46972 [Saccharomyces cerevisiae]"
join(<2185..2338,2432..2504,2922..3081,3204..>3281)
/gene="MZB10.1"
join(2185..2338,2432..2504,2922..3081,3204..3281)
/gene="MZB10.1"
/codon_start=1
/product="putative signal peptidase"
/protein_id="AAD56314.1"
/db_xref="GI:5923663"
/translation="MGIONLQVAKKSFSGSIIGLTISDCCSVYPRDGSMPFN
PQRNSYIDYVLDKFCILDYKFRAGDYYVPSPHFDRTIKRIVGMPGWISSSD
VIRPECHWCWEDNKTSLDSKSGPIPLGLIGQRTVRYWMPORISKIGR"
complement(3284..3360)
/note="exon predicted by xgrail, quality
marginal.shadowexon"
complement(3372..3486)
/note="exon predicted by xgrail, quality
marginal.shadowexon"
complement(3872..3962)
/note="exon predicted by xgrail, quality
marginal.shadowexon"
complement(3922..4055)
/rpt_family="(GAAA)n"
<4496..>5463
/gene="MZB10.2"
/note="similar to Yippee putative zinc binding protein
join(<4496..4591,5070..5113,5199..5307,5386..>5463)
/gene="MZB10.2"
join(4496..4591,5070..5113,5199..5307,5386..5463)
/gene="MZB10.2"
/codon_start=1
/product="Yippee-like protein"
/protein_id="AAD56315.1"
/db_xref="GI:5923664"
/translation="MKRLPVLDLGLVYSCRCYCTHPAVNDIISKSFHCKRRAYLF
DNVAVTVEKRRVYITGMVADIFCVSGSLVGWKYFIADKSKTKEGFTILR
FSSFNV"
complement(5490..5586)
/note="exon predicted by xgrail, quality
marginal.shadowexon"
5534..5615
misc_feature
/note="exon predicted by xgrail, quality excellent"
complement(5641..5725)
/note="exon predicted by xgrail, quality
excellent.shadowexon"
complement(6084..6095)
/note="exon predicted by xgrail, quality
marginal.shadowexon"
6193..6476
misc_feature
/note="exon predicted by xgrail, quality excellent"
complement(6418..6474)
/note="exon predicted by xgrail, quality
excellent.shadowexon"
complement(6632..6705)
/rpt_family="(GA)n"
<6920..>9232
/gene="MZB10.3"
/note="similar to unknown protein GB:AAD595292 [Arabidopsis
thaliana]"
join(<6920..7010,7097..7253,7541..7646,7762..8769,
8969..>9232)
/gene="MZB10.3"
join(6920..7010,7097..7253,7541..7646,7762..8769,
8969..9232)
/gene="MZB10.3"
/note="hypothetical protein"
/codon_start=1

/protein_id="A056316.1"
/db_xref="GI:5923665"
/translation="MLTHRDEELSLFLEMRBREKEHRADSLITGSDNSINATITAA
AAAALSGVSTFASSORPLRRTAAENLTVSENEKSDYDLITTPGTQPEKESRSTMA
NOHDAPSRFTVLKSLRGNCREDIVSGNNKPOTSSSVAGLRPSSGSSRSSTRA
TPRRSTPTPTSTSRPTVTRASNSRSTPSRALTAARATSTAAARRTSSSTSGAR
SAPTRSNPPRSSASKKPVSRPATPRRSTPGTIVSKAPSRSTSPVYNSIS
KAPRGTSPSPPLNSRPMKPEMPGSLAPPLRTTLADRPVSAERGLGVAASAG
SRSGTERRGGPTSGGNAAROSCSRRAPIGNGTSLGVREGAKASNGSGCD
NLSPVANGNKMYRYVMARKLGPRLLENKGRSGSKSSAFNSLGIQRNLSIDMA
IRHMDIRKGTGLRPLVTKVPASMTSVSRPSVSSSPVATSTVSSDSPVDNIN
ILCIDGAEANDLLSERSYAASPRNOYPRFTS"
/misc_feature
/note="exon predicted by xgrall, quality
marginal.shadowexon"
10828..10857
/note="exon predicted by xgrall, quality marginal"
/gene="MZB10.4"
/note="similar to receptor kinase GB:S70769 [Arabidopsis
thaliana]"
join(<11221..11263,11488..11600,11690..11900,11982..12219,
12298..12448,12597..>12992)
/gene="MZB10.4"
join(11221..11263,11488..11600,11690..11900,11982..12219,
12298..12448,12597..12992)
/gene="MZB10.4"
/codon_start=1
/product="putative receptor ser/thr protein kinase"
/protein_id="A056317.1"
/db_xref="GI:5923665"
/translation="MCGNSDRIGOREAEICTNNVRFVSNLSRATDSFPHNRIGG
GGYGVKGYLRDGTQYAVKSLSAESKQGREPTEINLSNTHHPVLKICCTIG
NNILVTEYELNNSLASVLGSKRYPLDMSKRAALCVSTAGSLAEVHEHVHA
RDIKASNILLDSNFSKPIDGLAKLPPDMVTHVSTGLAEVLAELVLLQGTTH
ADYSPFELILEVYISGNSSTRAAFGDEYVLEVMVLRERRLIEVDDELTFPDD
EYRFILVALFCTOAAOKRPNKOVEMLRKELNENALFEPGVYRDVGNRNRH
GIGLRVGOGSSGSESTORDKSSAVPQSGTNSISISITELAPR"
<14213..>15448
/gene="MZB10.5"
/note="similar to hypothetical protein GB:AAC27164
Query Match 100.0%; Score 2160; DB 8; Length 85561;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2160; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

421 TTGAAGAAACAGAGAAAGATGTTCTTGAACCTTACATCTCCAGGATATGCAAT 480
|||||
64744 TTGAAGAAACAGAGAAAGATGTTCTTGAACCTTACATCTCCAGGATATGCAAT 64803
Db
481 TACATTTGTTAGTGTGTGGAGAGATTAATTTCTCATGTTCTTGTCTGTGAATTTTG 540
|||||
64804 TACATTTGTTAGTGTGTGGAGAGATTAATTTCTCATGTTCTTGTCTGTGAATTTTG 64863
Db
541 GGTAAATGATTGATGTTGTCATTAGAACCAAAATAACTTTACTGTTATGACTGC 600
|||||
64864 GGTAAATGATTGATGTTGTCATTAGAACCAAAATAACTTTACTGTTATGACTGC 64923
Db
601 TTATATAGTAAAGTTACAGATTTGTTTTCTATACAGCAATGTTTCGAAAGTA 660
|||||
64924 TTATATAGTAAAGTTACAGATTTGTTTTCTATACAGCAATGTTTCGAAAGTA 64983
Db
661 TAGTAGCATGACAGACTAGATGTAGAGACCAATCAATGCTGCTTTAATCTTC 720
|||||
64984 TAGTAGCATGACAGACTAGATGTAGAGACCAATCAATGCTGCTTTAATCTTC 65043
Db
721 AAAGATTAGACTCTTGAGAGATATGCAATAGAGAAATGACAAATGCACTTCACTCA 780
|||||
65044 AAAGATTAGACTCTTGAGAGATATGCAATAGAGAAATGACAAATGCACTTCACTCA 65103
Db
781 GCAGTACAACTGATGCTTGTAGTCCATTTCTTCAATGATGATGTTGTTGTA 840
|||||
65104 GCAGTACAACTGATGCTTGTAGTCCATTTCTTCAATGATGATGTTGTTGTA 65163
Db
841 TCTTCGCTTAGTCTTTGACTTTCGCGGTTTCATGAGCCGATGATGATCTCCATAT 900
|||||
65164 TCTTCGCTTAGTCTTTGACTTTCGCGGTTTCATGAGCCGATGATGATCTCCATAT 65223
Db
901 TCTGTGCTAACTTTTACATCTACCAACGTTAAATAGTTGTAATGTAATCTTCT 960
|||||
65224 TCTGTGCTAACTTTTACATCTACCAACGTTAAATAGTTGTAATGTAATCTTCT 65283
Db
961 AGTATGCTGAGTATACAGTGTCTGTATGTGCCAAACGATGTCATATTTACTGTAT 1020
|||||
65284 AGTATGCTGAGTATACAGTGTCTGTATGTGCCAAACGATGTCATATTTACTGTAT 65343
Db
1021 GTTCTTCTTTAGGACCTTAATCTTTCATCACTGACAGCAGAGATTTACCAAGA 1080
|||||
65344 GTTCTTCTTTAGGACCTTAATCTTTCATCACTGACAGCAGAGATTTACCAAGA 65403
Db
1081 CAAGTATTTATACAAAGATATGTCATATATACAAATATGCTGAGTACCAAGAT 1140
|||||
65404 CAAGTATTTATACAAAGATATGTCATATATACAAATATGCTGAGTACCAAGAT 65463
Db
1141 ACACCAATTTACCAATTTGCAAGTAACCTAATGCTGAGTAAATGACTGATCTGAG 1200
|||||
65464 ACACCAATTTACCAATTTGCAAGTAACCTAATGCTGAGTAAATGACTGATCTGAG 65523
Db
1201 ATTATATTTGCAAGCTTTCCATGAGGAGGAAATTTGACTGTAATCCATTAAGTTTTC 1260
|||||
65524 ATTATATTTGCAAGCTTTCCATGAGGAGGAAATTTGACTGTAATCCATTAAGTTTTC 65583
Db
1261 TCGCCTTTGGTGTGAGTCTAGGTTTTCGTAGCAAAAGAAACATTAAGCTTTGTTTC 1320
|||||
65584 TCGCCTTTGGTGTGAGTCTAGGTTTTCGTAGCAAAAGAAACATTAAGCTTTGTTTC 65643
Db
1321 TCGCCTTTGTAGAGTACTATCAGGTATATCTCAGGCGCCAAAGCTTAAGTTTATG 1380
|||||
65644 TCGCCTTTGTAGAGTACTATCAGGTATATCTCAGGCGCCAAAGCTTAAGTTTATG 65703
Db
1381 GAAACTTTGACTGGAATCTATCTCTTCTCTACAGGACTGCTTGAGATGCAATTC 1440
|||||
65704 GAAACTTTGACTGGAATCTATCTCTCTCTCTACAGGACTGCTTGAGATGCAATTC 65763
Db
1441 AAGTGAAGGAGAGATGGAACATCTATCTGAGAGCCAAATTTGGAAGCACACAAAGT 1500
|||||
65764 AAGTGAAGGAGAGATGGAACATCTATCTGAGAGCCAAATTTGGAAGCACACAAAGT 65823
Db

```

Oy 1501 ACCGACATGGCAGACACAAAGGATTCATTCCTTTGTGTAATTCATGATTTGA 1560
    |||||||
Db 65824 ACCGTGACATGGCAGACACAAAGGATTCATTCCTTTGTGTAATTCATGATTTGA 65883
Oy 1561 ACAGACATGATGCTGCTGTTGTACATGATGATTTGATTTGGTGGTGGCAGATCCTGGA 1620
    |||||||
Db 65884 ACAGACATGATGCTGCTGTTGTACATGATGATTTGATTTGGTGGTGGCAGATCCTGGA 65943
Oy 1621 CATGCTCTTTTAAAGATTTAGTGGTGAACCAAGCAAGGATTAATAAATTTGATTC 1680
    |||||||
Db 65944 CATGCTCTTTTAAAGATTTAGTGGTGAACCAAGCAAGGATTAATAAATTTGATTC 66003
Oy 1681 CATGATGCTCCCATTTATTAATTAAGCTTGTAAGATGTTGAAATGATTTGACAAATC 1740
    |||||||
Db 66004 CATGATGCTCCCATTTATTAATTAAGCTTGTAAGATGTTGAAATGATTTGACAAATC 66063
Oy 1741 AGGAGCTGCTAAGGATTTCTGTTCAATGGGGTGGATGAGTTAGGCACAAACATTTCA 1800
    |||||||
Db 66064 AGGAGCTGCTAAGGATTTCTGTTCAATGGGGTGGATGAGTTAGGCACAAACATTTCA 66123
Oy 1801 TTGATTAATTTCCAGAGTACCAACAGAGATGAGTGTGAAGCATTAACAGATTTCA 1860
    |||||||
Db 66124 TTGATTAATTTCCAGAGTACCAACAGAGATGAGTGTGAAGCATTAACAGATTTCA 66183
Oy 1861 TTGGGAGTCAATGAAAGCTGCTGATGGGATTTAACAAGCAATTTATCGGCTAACAT 1920
    |||||||
Db 66184 TTGGGAGTCAATGAAAGCTGCTGATGGGATTTAACAAGCAATTTATCGGCTAACAT 66243
Oy 1921 GATATATGTGAAAGCTGATTTCCAGAGTCAATCAACAGAGAGAGTGAAGTAACTAGGA 1980
    |||||||
Db 66244 GATATATGTGAAAGCTGATTTCCAGAGTCAATCAACAGAGAGAGTGAAGTAACTAGGA 66303
Oy 1981 AGTAGGAATTAAGAAAGCAGCATGAGAGAGTCTCTCAGGCTATCTCATTCTCAAGATGA 2040
    |||||||
Db 66304 AGTAGGAATTAAGAAAGCAGCATGAGAGAGTCTCTCAGGCTATCTCATTCTCAAGATGA 66363
Oy 2041 TTGTTAGTATACCATCATTTGATTTTACAAAGTGTGATCTCAGCCCTTCATCAAAAT 2100
    |||||||
Db 66364 TTGTTAGTATACCATCATTTGATTTTACAAAGTGTGATCTCAGCCCTTCATCAAAAT 66423
Oy 2101 GAGATATCCTCGAGTATGATGATTTTAAATGAATAATGATTCGCTCTACCTATCAACA 2160
    |||||||
Db 66424 GAGATATCCTCGAGTATGATGATTTTAAATGAATAATGATTCGCTCTACCTATCAACA 66483

```

```

RESULT 2
ATAC011436 106688 bp DNA linear PLN 24-JAN-2001
LOCUS Arabidopsis thaliana chromosome III BAC F3L24 genomic sequence,
DEFINITION complete sequence.
ACCESSION AC011436
VERSION AC011436.7 GI:12408733
KEYWORDS HTG.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
REFERENCE 1 (bases 1 to 106688)
AUTHORS Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.,
Bowman,C.M., Koo,H., Fujii,C.Y., Uteback,T.R., Barnstead,M.E.,
Rommang,C.L., White,O., Nielsen,W.C. and Fraser,C.M.
TITLE Arabidopsis thaliana chromosome III BAC F3L24 genomic sequence
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 106688)
Lin,X. and Kaul,S.
TITLE Direct Submission
JOURNAL Submitted (06-OCT-1999) The Institute for Genomic Research, 9712
Medical Center Dr. Rockville, MD 20850, USA, xlin@tigr.org
REFERENCE 3 (bases 1 to 106688)
AUTHORS Lin,X.
TITLE Direct Submission
JOURNAL Submitted (24-JAN-2001) The Institute for Genomic Research, 9712

```

COMMENT

Medical Center Dr., Rockville, MD 20850, USA
 On Jan 24, 2001 this sequence version replaced gi:12280819.
 Address all correspondence to:

Xiaoying Lin
 The Institute for Genomic Research
 9712 Medical Center Dr.
 Rockville, MD 20850, USA
 e-mail: xlin@tigr.org

BAC clone F3L24 is from Arabidopsis chromosome III and is near the
 molecular marker C1C7A12.
 The orientation of the sequence is from SP6 to T7 end of the BAC
 clone.

Genes were identified by a combination of three methods: Gene
 prediction programs including GNAI (available by anonymous ftp
 from arthur.epm.ornl.gov), GeneIndex (Phil Green, University of
 Washington), GenScan (Chris Burge,
<http://www.cbs.dtu.dk/netgene/cbsnetgene.html>), searches of the
 complete sequence against a peptide database and the Arabidopsis
 EST database at TIGR (<http://www.tigr.org/tdb/at/est.html>).
 Annotated genes are named to indicate the level of evidence for
 their annotation. Genes with similarity to other proteins are named
 after the database hits. Genes without significant peptide
 similarity but with EST similarity are named as 'unknown' proteins.
 Genes without protein or EST similarity, that are predicted by more
 than two gene prediction programs over most of their length are
 annotated as 'hypothetical' proteins. Genes encoding tRNAs are
 predicted by tRNAscan-SE (Sean Eddy,
<http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are
 identified by RepeatMasker (Arian Smit,
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Regions of
 genomic sequence that are not annotated as genes but have predicted
 exons by GNAI are annotated as misc features.

FEATURES

source

1. 106688
 /organism="Arabidopsis thaliana,"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="III"
 /map="C1C7A12"
 /clone="F3L24"
 1. 21238
 /note="overlap with BAC clone MZB10
 (AC009326:64324..85561)."

misc_feature

1. 122
 /note="exon predicted by xgrail, quality
 marginal, shadowexon"
 248..>1916
 /gene="F3L24.1"
 /note="MZB10.18; nearly identical to hypothetical protein
 GB:AAD56331 [Arabidopsis thaliana]"
 join<248..469,653..769,852..946,1033..1126,1213..1344,
 1419..1523,1612..1662,1743..>1916)
 /gene="F3L24.1"
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="AAFI017.1"
 /db_xref="GI:6478912"

gene

1. 1916
 /translation="WALSMERFSGISGCFKAPNPVLISAPNKINFTLRRKRRL
 RVSAYSEKFAESALETRKRIVLEPSLQKYSMTLQDTELOMLAFSSKIRLL
 RSMALNETMOVFDPAGMEPEYDFICAGNEFTSTNIVLIDNPHLOLDIDVO
 DRYVNRKMSIYHKAETFPWGKLGESIKFEPSPLVMYTRSSSEKKAFAFSALEY
 YQAMLEMTIOYREMEPSHVANCAQKITYTRAKDPGGLKRLVGEAKAEELLR
 DFLRNQDELGKTFIDTFPEYQEDGTVSDKRSITGSKSYETRPWLDGPFG"
 complement(1934..2024)
 /note="exon predicted by xgrail, quality
 marginal, shadowexon"
 /gene="F3L24.2"
 /note="MZB10.19; identical to hypothetical protein

CDS

misc_feature

gene

GB:AD56336 [Arabidopsis thaliana], contains Pfam profile: RNA recognition motif (aka RRM, RBD, or RNP domain)"
complement(join(<2362, .2462,2534, .2703,3004, .3276, 3353, .3384,3475, .>34951))
/gene="F3L24.2"
complement(join(2362, .2462,2534, .2703,3004, .3276, 3353, .3384,3475, .34951))
/gene="F3L24.2"
/codon_start=1
/product="putative RNA-binding protein"
/protein_id="AAE14010.1"
/db_xref="GI:6478935"
/translation="MDEIANKKLEVKRSKRIKDSIDIERISVTGYDIELPREHVESELEK
KLEFSGGITYVLEIDETPNSLSMSAFYIPEVGEAGADRALQNGSDMGWTAQDAEDP
EPKEDNNSIYVLEIDEDIDKPKMTSGDGMGGRAVFAVFGHNRNIFPHRPGRI
ICGRDEKPKMAIKIKIKPKPRKAKRIKKPKMTLE"
4139, .4182
/note="exon predicted by xgrail, quality
marginal_shadowexon"
complement(4387, .4429)
/note="exon predicted by xgrail, quality good_shadowexon"
complement(4553, .4736)
/note="exon predicted by xgrail, quality good_shadowexon"
5389,5442
/note="exon predicted by xgrail, quality good"
5748, .9002
/gene="F3L24.3"
/note="MZB10.20; similar to MURA transposase of maize
Mutator transposon"
/pseudo
9051, .9107
/note="exon predicted by xgrail, quality good"
complement(9181, .9241)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
9601,9827
/note="exon predicted by xgrail, quality excellent"
complement(9619, .9780)
/note="exon predicted by xgrail, quality good_shadowexon"
complement(<10632, .>14351)
/gene="F3L24.4"
/note="MZB10.21; identical to hypothetical protein
GB:AD56332 [Arabidopsis thaliana]"
complement(join(<10633, .10789,10875, .11340,11454, .11538
11622, .12016,12222, .12341,12417, .12578,12644, .13375,
13467, .>14351))
/gene="F3L24.4"
complement(join(10632, .10789,10875, .11340,11454, .11538,
11622, .12016,12222, .12341,12417, .12578,12644, .13375,
13467, .>14351))
/codon_start=1
/protein_id="AAE14010.1"
/db_xref="GI:6478913"
/translation="MGDELPRLEFRLGSEPRGRKRNINFTLRWIEIKALDONLH
KOLSGSGFLKMOGTMOFSVFLHYILSRQVLPKTYEMMWFAKPIRFAIDPRL
VTGLSGSGFLKMOGTMOFSVFLHYILSRQVLPKTYEMMWFAKPIRFAIDPRL
DLTPRFLSTLLNGCIIKPCPSRPIIDPDHAKMDVSEFKIYPMGRSPLATIRSA
KORTAKELADQTTALOGCPHPLAVLVACAPATITFASVFIIDPITPEVLELVTVL
GLSLINVNDATIDOGSDMRVTSILNLELAKELIDLSDEDEDEEYSLIMEE
AGEFVINSNGGNDADVHSPDEDFMADCVNOAGDGDDEMDVDSGVAAEERECIA
HGEVFDGADGEGENADVPYDKGVSFADDEGLQOVPRAKGEENNDGSGAAGLEAV
IKEYVDEKEERLKIIOGNNBELKVSFAEYIKKELVAILTESVREAVSLNGSG
ITVSGCKVADAAHSQVYKHNILVETSSALDQITADQPGKHKHSQOYEVNIGSEG
SDNLPAACGIVNGSVPEKSNNTMEVREEDSIQKGRKKSQSDQASNPPLKRAK
DMEHVAIRSYVSPERLIDEDVIVPNSFLGTOEINDHRSRLPRGKKRTPSKVY
LRSNHRIDRPLPSSINDNMSIIPQPHOTASRGAKSHVSSSAKEKTPMOSPFI
GPFKEFVSPESSVYKFLKMOGRSVYLVADVLEWFEFLPIFGQITLDSGPDLL
VRFIOTLIDKKKKVLEFIDLPSSELSLRQYKPYGADKSAFYESTLARVPMVPRK
WFAEDLVITCQNVGRQHWGLVYIDLKKKGVTYLDQNRDYLDEQVEKVEHITMLP
LIDRKHGNGVGLVLETPMVSRRKIIOGSCSGLVGLASVLLLEHVSVMMEYCGK
LIDGKMDAAKQVLALETNALSPTRGCA"

	misc-feature	14460..14526	/note="exon predicted by xgrall, quality excellent-shadowexon"
	misc-feature	15070..15229	/note="exon predicted by xgrall, quality marginal"
	gene	15834..>17202	/gene="F3I24.5"
		/note="MZB10.22; identical to hypothetical protein gb:AAD5633 [Arabidopsis thaliana]"	
	mRNA	join(15834..16133,16228..16393,16566..>17202)	
		/gene="F3I24.5"	
	CDS	join(15860..16133,16228..16393,16566..17202)	
		/gene="F3I24.5"	
		/note="unknown protein"	
		/codon_start=1	
		/protein_id="AAF14019.1"	
		/db_xref="GI:6478914"	
		/translation="MOTLHQSQLONPRAANNSESAPKQYAOAMERLNQAARYADILRGSRILEAMFVASOPDRHTMPQLFLREASMKOHLDRLIGKLTSEESGLTLESLSRSNMLHPPLVCPDAVVAYAMKRLOAQAGSAVDRLRLAKPTDGRRRFPNIIDDGKLMPSSSKRHASHLLLENGESEVVDKTEPDIOISAEIKLVPSKYTGKLNWIKRANSLPLSGSDDPTEAEVKRIPFOSSSKRSGLQEDVEVDKLAIEVLSPSIFRALVSLEPAGSVDPDAVAFPSPDVTSNLALKTDBSHVLFKFSCESTFFLOGSYLHNH RGESEVHYHKITVIKKMLIIIV"	
	misc-feature	17250..17318	
		/note="exon predicted by xgrall, quality excellent"	
	misc-feature	17412..17553	
		/note="exon predicted by xgrall, quality excellent"	
	Query Match	100.0%; Score 2160; DB 8; Length 106688;	
	Best Local Similarity	100.0%; Pred. No. 0;	
	Matches 2160; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
Oy	1	GAATTCGCCAGCTCAACGTGCATCGCATCCAGTGGCGGATGNGGCCCTATAGTTGG	60
Dd	1	GAAATCCCCACGTCAACGTGCATCGCATCCAGTGGCGGATGNGGCCCTATAGTTGG	60
Oy	61	ACCATTGACTCGGAGGAGTGTGAAATTCATTGTCGTTGCCAATTCGTTGTCTCCTAC	120
Dd	61	ACCATTGACTCGGAGGAGTGTGAAATTCATTGTCGTTGCCAATTCGTTGTCTCCTAC	120
Oy	121	AACGTGTGAAAAATTTTATCTCTTTATATAGATAAAGATCTGCTTTTTCACTTTTCAGTA	180
Dd	121	AACGTGTGAAAAATTTTATCTCTTTATATAGATAAAGATCTGCTTTTTCACTTTTCAGTA	180
Oy	181	TGAAGAANAATGGAAGAGAGTCGCGAGAGAAGAGAACCTTTGCGTTTCAGTTTGAGAGCT	240
Dd	181	TGAAGAANAATGGAAGAGAGTCGCGAGAGAAGAGAACCTTTGCGTTTCAGTTTGAGAGCT	240
Oy	241	TGTTGTATAGCTTTATCAATGAGATTTGGGTTTTCATTTGGGTGATGCTTCAAGCAC	300
Dd	241	TGTTGTATAGCTTTATCAATGAGATTTGGGTTTTCATTTGGGTGATGCTTCAAGCAC	300
Oy	301	AAACCCACCTGTTCTAATCTCTGCAAGCCCTPATATAGATCAATTTACGTTGAGAGAG	360
Dd	301	AAACCCACCTGTTCTAATCTCTGCAAGCCCTPATATAGATCAATTTACGTTGAGAGAG	360
Oy	361	AAAAGAAAAGATTTCTACTTAGAGTCCTGCTGTCGATATAGGAATTCGAGAGCTGCG	420
Dd	361	AAAAGAAAAGATTTCTACTTAGAGTCCTGCTGTCGATATAGGAATTCGAGAGCTGCG	420
Oy	421	TTTGTAGAGAAACCGAGAAAGATCGTTCTTGAACCTTCACATGCTCCAGATATATGCAAT	480
Dd	421	TTTGTAGAGAAACCGAGAAAGATCGTTCTTGAACCTTCACATGCTCCAGATATATGCAAT	480
Oy	481	TACATTTTCGTTAGTGTAGTGGAGGATTATATTTCTCAATGTTTCTTGCTGTGAATTTTG	540
Dd	481	TACATTTTCGTTAGTGTAGTGGAGGATTATATTTCTCAATGTTTCTTGCTGTGAATTTTG	540
Oy	541	GGTAAATTTGATTTGATTTGCTATTAGAAACCAACAATACTTACTGTTATATGACTGC	600
Dd	541	GGTAAATTTGATTTGATTTGCTATTAGAAACCAACAATACTTACTGTTATATGACTGC	600

Db	1681	CATTAGTGTGCCCATTAATTAATTAACCTTGTAAGATGTTGAAGAATTTGGACAATAATC	1740
OY	1741	AGGACTGCTGAAGGATTTTCCTGTTCAATGGGGTGATGAGTAGGCACAAAACATTCA	1800
Db	1741	AGGACTGCTGAAGGATTTTCCTGTTCAATGGGGTGATGAGTAGGCACAAAACATTCA	1800
OY	1801	TTCATTACTTTCCACAGATACCACAGAAAGATGGAACGTGAAGGATTAAGGAAGTATCA	1860
Db	1801	TTCATTACTTTCCACAGATACCACAGAAAGATGGAACGTGAAGGATTAAGGAAGTATCA	1860
OY	1861	TTGGGAAGTCAATGAAACTCGTCOCATGGGATTTAACAGACAAATTTATCGGCTAACAT	1920
Db	1861	TTGGGAAGTCAATGAAACTCGTCOCATGGGATTTAACAGACAAATTTATCGGCTAACAT	1920
OY	1921	GATATATGTGAACAGATGAGATTTTCAGAGTCATCAACACAGAGAGCTGAACCTTAGGA	1980
Db	1921	GATATATGTGAACAGATGAGATTTTCAGAGTCATCAACACAGAGAGCTGAACCTTAGGA	1980
OY	1981	AGTAGAATAAGAAAGAGCACAGATGAGAGTCTCTCAGGTCTATCTCATTTCAAGATGA	2040
Db	1981	AGTAGAATAAGAAAGAGCACAGATGAGAGTCTCTCAGGTCTATCTCATTTCAAGATGA	2040
OY	2041	TTGTTTGAAGTACCATGATGATTTGATTTTACAGTGAAGTCTCAGCCCTTCATCAAAAT	2100
Db	2041	TTGTTTGAAGTACCATGATGATTTGATTTTACAGTGAAGTCTCAGCCCTTCATCAAAAT	2100
OY	2101	GAGAAATCCCTGAGATGATGATTTTAATGAAGAAATATATTCGCTCTACTACTATACAGA	2160
Db	2101	GAGAAATCCCTGAGATGATGATTTTAATGAAGAAATATATTCGCTCTACTACTATACAGA	2160
RESULT 3 AY099706			
LOCUS			
DEFINITION	Arabidopsis thaliana unknown protein (At3g09150) mRNA, complete cds.		
ACCESSION	AY099706	1358 bp	mRNA linear PLN 06-MAY-2002
VERSION	AY099706.1	GI:20466479	
KEYWORDS	FLI CDNA.		
SOURCE	Arabidopsis thaliana.		
ORGANISM	Arabidopsis thaliana. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopses. 1 (bases 1 to 1358)		
REFERENCE	Nguyen,M., Karlin-Neumann,G., Southwick,A., Lam,B., Miranda,M., Palm,C.J., Bower,L., Jones,T., Banh,J., Carninci,P., Chen,H., Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamiya,A., Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K., Sakano,H., Sakurai,T., Satou,M., Seki,M., Shimn,P., Yamada,K., Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W. Direct Submission		
AUTHORS	Submitted (24-APR-2002) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		
TITLE	e-mail for correspondence: arab@sequence.stanford.edu		
JOURNAL			
COMMENT			
	RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : "RIKEN Arabidopsis Full-length cDNA") : Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.		
	The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Nguyen,M., Southwick,A., Karlin-Neumann,G., Lam,B., Miranda,M., Palm,C.J., Bower,L., Jones,T., Banh,J., Chen,H., Cheuk,R., Chung,M.K., Kim,C., Lin,J., Liu,S.X., Pham,P.K., Sakano,H., Shimn,P., Yamada,K., Ecker,J., Theologis,A. and Davis,R.W.		
	Nguyen,M. (SSF/Stanford) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Davis P.W.		

(SSP/Stanford) contributed equally to this work as PIs.

FEATURES
source
1. 1358
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="3"
/clone="RAF109-83-B22 (R25336)"
/note="This clone is in pBluescript
ecotype: Columbia"
1. 1358
/gene="At3g09150"
/note="F3L24.1"
139. 1128
/gene="At3g09150"
/codon_start=1
/product="unknown protein"
/protein_id="PAM2057.1"
/db_xref="GI:20466480"
/translation="MALSMERFSGSCFKAPNPVLISAPNKINFLRRKKRFL
RVSASVKEFAESALERTRRIVLEPRLQEKYSMTGLDQTELOMLAFKSRIRL
RSMALNEMTMOVDFAGMEPEYDPTFCANFPTSTVNVIVLDNPLHOLDOTDIO
DKYNNKIMSTYHKIAETFPWGKLTGSIKFSEFLVMTTRSSSKERKALFSALEY
YQALNEMTIQVREMEPSHVANCAEQHKYLTWRQKDPGHLKRLVGEKAKELLR
DFLNGVDELGTFTFDYFPEYQTEDETVSDKRSIIKSKSYETRPWDLTGQFIG"
BASE COUNT 426 a 231 c 308 g 393 t
ORIGIN
Query Match 18.8%; Score 406; DB 8; Length 1358;
Best Local Similarity 100.0%; Pred. No. 1e-206;
Matches 406; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1741 AGGAGCTGCTAAGGATTTCTGTTCAATGGGGTGATGATGAGCAAAAACATTCA 1800
DB 953 AGGAGCTGCTAAGGATTTCTGTTCAATGGGGTGATGATGAGCAAAAACATTCA 1012
QY 1801 TTGATTACTTCCAGAGTACCAACAGAGATGAGACTGTAAGCGATAACGAATATCA 1860
DB 1013 TTGATTACTTCCAGAGTACCAACAGAGATGAGACTGTAAGCGATAACGAATATCA 1072
QY 1861 TTGGGAGTCAATATGAACACTGTCATGGGATTTAACAAGCAATTTATCGGCTAACAT 1920
DB 1073 TTGGGAGTCAATATGAACACTGTCATGGGATTTAACAAGCAATTTATCGGCTAACAT 1132
QY 1921 GATATATGTGAACAAGTCAAGTTCAGAGTCATCAACAAGAGAGAGTGAACCTTAGGGA 1980
DB 1133 GATATATGTGAACAAGTTCAGAGTCATCAACAAGAGAGAGTGAACCTTAGGGA 1192
QY 1981 AGTAGAATAAGAAAGAGAGAGAGAGTCTCAGAGTCTATCTGATTTCAAGATGA 2040
DB 1193 AGTAGAATAAGAAAGAGAGAGAGTCTCAGAGTCTATCTGATTTCAAGATGA 1252
QY 2041 TTGTTTAGTACATGATTTGTAGTTTACAACTGTAAGTCTCAGAGCTTCATCAAAAT 2100
DB 1253 TTGTTTAGTACATGATTTGTAGTTTACAACTGTAAGTCTCAGAGCTTCATCAAAAT 1312
QY 2101 GAGATATCTCGAGATGATATGATTTAATGAATGATATCGTCT 2146
DB 1313 GAGATATCTCGAGATGATATGATTTAATGAATGATATCGTCT 1358
RESULT 4
AB045112 1337 bp mRNA linear PLN 03-MAR-2001
LOCUS Arabidopsis thaliana mRNA for phytochromobilin synthase HY2
DEFINITION protein, complete cds.
ACCESSION AB045112
VERSION AB045112.1 GI:13359272
KEYWORDS
SOURCE Arabidopsis thaliana cDNA to mRNA.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (sites)
AUTHORS Kohchi,T., Mukougawa,K., Frankenberg,N., Masuda,M., Yokota,A. and
Lagaras,J.C.
TITLE The arabidopsis hy2 gene encodes phytochromobilin synthase, a
ferredoxin-dependent biliverdin reductase
JOURNAL Plant Cell 13 (2), 425-436 (2001)
MEDLINE 21124703
REFERENCE 2 (bases 1 to 1337)
AUTHORS Kohchi,T.
TITLE Direct Submission
JOURNAL Submitted (21-JUN-2000) Takayuki Kohchi, Nara Institute of Science
and Technology, Graduate School of Biological Sciences; 8916-5
Takayama, Ikoma, Nara 630-0101, Japan
(E-mail:kouchi@nara.ac.jp, Tel:81-743-72-5561,
Fax:81-743-72-5569)
FEATURES
source
Location/Qualifiers
1. 1337
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
1. 1337
/gene="HY2"
1. 95
/gene="HY2"
96. 1085
/gene="HY2"
/codon_start=1
/product="phytochromobilin synthase HY2 protein"
/protein_id="BAB3374.1"
/db_xref="GI:13359273"
/translation="MALSMERFSGSCFKAPNPVLISAPNKINFLRRKKRFL
RVSASVKEFAESALERTRRIVLEPRLQEKYSMTGLDQTELOMLAFKSRIRL
RSMALNEMTMOVDFAGMEPEYDPTFCANFPTSTVNVIVLDNPLHOLDOTDIO
DKYNNKIMSTYHKIAETFPWGKLTGSIKFSEFLVMTTRSSSKERKALFSALEY
YQALNEMTIQVREMEPSHVANCAEQHKYLTWRQKDPGHLKRLVGEKAKELLR
DFLNGVDELGTFTFDYFPEYQTEDETVSDKRSIIKSKSYETRPWDLTGQFIG"
3'UTR
polyA_site
BASE COUNT 434 a 226 c 301 g 376 t
ORIGIN
Query Match 17.6%; Score 381; DB 8; Length 1337;
Best Local Similarity 100.0%; Pred. No. 3e-193;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1741 AGGAGCTGCTAAGGATTTCTGTTCAATGGGGTGATGATGAGCAAAAACATTCA 1800
DB 910 AGGAGCTGCTAAGGATTTCTGTTCAATGGGGTGATGATGAGCAAAAACATTCA 969
QY 1801 TTGATTACTTCCAGAGTACCAACAGAGATGAGACTGTAAGCGATAACGAATATCA 1860
DB 970 TTGATTACTTCCAGAGTACCAACAGAGATGAGACTGTAAGCGATAACGAATATCA 1029
QY 1861 TTGGGAGTCAATATGAACACTGTCATGGGATTTAACAAGCAATTTATCGGCTAACAT 1920
DB 1030 TTGGGAGTCAATATGAACACTGTCATGGGATTTAACAAGCAATTTATCGGCTAACAT 1089
QY 1921 GATATATGTGAACAAGTTCAGAGTCATCAACAAGAGAGAGTGAACCTTAGGGA 1980
DB 1090 GATATATGTGAACAAGTTCAGAGTCATCAACAAGAGAGAGTGAACCTTAGGGA 1149
QY 1981 AGTAGAATAAGAAAGAGAGAGAGTCTCAGAGTCTATCTGATTTCAAGATGA 2040
DB 1150 AGTAGAATAAGAAAGAGAGAGAGTCTCAGAGTCTATCTGATTTCAAGATGA 1209
QY 2041 TTGTTTAGTACATGATTTGTAGTTTACAACTGTAAGTCTCAGAGCTTCATCAAAAT 2100
DB 1210 TTGTTTAGTACATGATTTGTAGTTTACAACTGTAAGTCTCAGAGCTTCATCAAAAT 1269
QY 2101 GAGATATCTCGAGATGATATGATTTAATGAATGATATCGTCT 2146
DB 1313 GAGATATCTCGAGATGATATGATTTAATGAATGATATCGTCT 1358

```

Db      1270 GAGATCCTCGATGATATAT 1290

RESULT 5
LOCUS   AY088874
DEFINITION Arabidopsis thaliana clone 98026 mRNA, complete sequence.
ACCESSION AY088874
VERSION   AY088874.1 GI:21407648
KEYWORDS
SOURCE   FLI_CDNA.
ORGANISM Arabidopsis thaliana
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
          1 (bases 1 to 1308)
          Haas,B.J., Volkovskiy,N., Town,C.D., Troukhan,M., Alexandrov,N.,
          Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.
          Full-length messenger RNA sequences greatly improve genome
          annotation
REFERENCE
AUTHORS  Genome Biol. (2002) In press
          2 (bases 1 to 1308)
          Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
          Feldmann,K.
          Full-length cDNA from Arabidopsis thaliana
          unpublished
          3 (bases 1 to 1308)
          Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
          Feldmann,K.
          Direct Submission
          Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road,
          Malibu, CA 90265, USA
COMMENT  This clone sequence is one of 5,000 Ceres full-length cDNAs made
          available to TIGR and Genbank. The following quality assessment of
          this set was done by comparison with known proteins: two percent of
          the clones are estimated to be 5'-truncated; less than one percent
          are 3'-truncated; approximately two percent represent alternative
          splice variants, including unspliced introns and spliced exons; one
          percent may contain premature stop codons; five percent may have
          frame shifts in a coding region. A sequence is considered to be
          5'-truncated if it lacks the translation initiation start (ATG). A
          sequence is considered to be 3'-truncated if it lacks the
          C-terminal end of the encoded protein. Please note that these cDNA
          sequences are derived from the WS or Laer ecotypes and therefore
          may contain polymorphisms when compared to sequences from Col-0.
          Genest carried out the library production and sequencing of the
          full-length clones. Ceres, Inc. carried out the clustering of the
          5' sequences, selection of clones, and sequence assembly.
FEATURES
Source   Location/Qualifiers
          1..1308
          /organism="Arabidopsis thaliana"
          /db_xref="taxon:3702"
          /clone="98026"
          82..1053
          /product="unknown"
          /protein_id="AA067180.1"
          /db_xref="GI:21618130"
          /translation="MEGFSIGSCFKAPNPVLISASPNKINFTLRNRKFLRLYASA
          VSKREASALLETKRRIYEPHSHLOYSSTGIDGTELOMLAFKSKITLISMAIE
          NETWQVDFAGEWEPETDIPICANFTSTNVIIVDLNPLHLQDIDYQKYLK
          IMSIYHKAETFPWGLKIGESIKFSPILWMTFRSSSKRKHLSAFLELYQAMLE
          MTIOVHEMESHRANCAEOKHRYITWRRAQKDPGHGLKRLVEAAKELLRDFLNG
          VDEIGRTKTFIDYPEYOTEDGTVSDKRSITIGSKSYETRPMDLTGQFTIG"
BASE COUNT  414 a      226 c      296 g      372 t
ORIGIN
Query Match      14.7%; Score 318; DB 8; Length 1308;
Best Local Similarity 99.5%; Pred. No. 2.6e-159;
Matches 418; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
          1741 AGGAGCTGCTAAGGATTTCTTCATGGGCGATGCTTAGCACAACAAACATTC 1800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||

```

```

Db      878 AGGAGCTGCTAAGGATTTCTTCATGGGCGATGCTTAGCACAACAAACATTC 937
OY      1801 TTGATTACTTTCCAGAGTACCAAGAAAGATGGAAGTAAAGCATATATCA 1860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      938 TTGATTACTTTCCAGAGTACCAAGAAAGATGGAAGTAAAGCATATATCA 997
OY      1861 TTGGGAAGTCATATGAAATCGGCGATGATTTTAAAGCAGCAATTTATGCGCTAACAT 1920
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      998 TTGGGAAGTCATATGAAATCGGCGATGATTTTAAAGCAGCAATTTATGCGCTAACAT 1057
OY      1921 GATATATGTGAACAACTCAGATTTTCAGACTCATCAACAGAGAGCTGAATCTTAGGA 1980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1058 GATATATGTGAACAACTCAGATTTTCAGACTCATCAACAGAGAGCTGAATCTTAGGA 1117
OY      1981 AGTGAATAAAGAAAGAGAGCATGAGAGCTGCTCATGCTATGCAATTCAGATGA 2040
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1118 AGTGAATAAAGAAAGAGCATGAGAGCTGCTCATGCTATGCAATTCAGATGA 1177
OY      2041 TTGTTGAGTTACCATGATGATTTTATTAAGTGTACCTGCAAGCCCTTCATCAAAAT 2100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1178 TTGTTGAGTTACCATGATGATTTTATTAAGTGTACCTGCAAGCCCTTCATCAAAAT 1237
OY      2101 GAGATCCTCGAGTATGATGATTTTATTAAGTGTACCTGCTCTACTATCAATCA 2160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1238 GAGATCCTCGAGTATGATGATTTTATTAAGTGTACCTGCTCTACTATCAATCA 1297

RESULT 6
LOCUS   AY128900
DEFINITION Arabidopsis thaliana unknown protein (At3g09150) mRNA, complete cds.
ACCESSION AY128900
VERSION   AY128900.1 GI:22136443
KEYWORDS
ORGANISM Arabidopsis thaliana
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
          1 (bases 1 to 1092)
REFERENCE
AUTHORS  Tripp,M., Southwick,A., Karlin-Neumann,G., Nguyen,M., Miranda,M.,
          Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H.,
          Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamiya,A.,
          Kawai,J., Kim,C., Liu,J., Liu,S.X., Narusaka,M., Pham,P.K.,
          Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K.,
          Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.
          Direct Submission
          Submitted (01-JUL-2002) DNA Sequencing and Technology Center,
          Stanford University, 855 California Avenue, Palo Alto, CA 94304,
          USA
          e-mail for correspondence: arab@sequence.stanford.edu
COMMENT  The RIKEN Genomic Sciences Center (GSC) members carried out the
          collection and clustering of RAFI cDNAs (RAFI cDNA: 'RIKEN
          Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
          Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
          Hayashizaki,Y. and Shinozaki,K.
          The Salk, Stanford, PGEC (SSP) Consortium members constructed and
          sequenced the PUN1 (ORF) clones using the RAFI cDNAs: Tripp,M.,
          Nguyen,M., Southwick,A., Karlin-Neumann,G., Lam,B., Miranda,M.,
          Palm,C.J., Bowser,L., Jones,T., Banh,J., Chen,H., Cheuk,R.,
          Chung,M.K., Kim,C., Liu,J., Liu,S.X., Pham,P.K., Sakano,H.,
          Shinn,P., Yamada,K., Ecker,J., Theologis,A. and Davis,R.W.
          Tripp,M. (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed equally
          to this work. Shinozaki,K. (RIKEN GSC) and Davis,R.W.
          (SSP/Stanford) contributed equally to this work as PIs.
          Location/Qualifiers
          1..1092
          /organism="Arabidopsis thaliana"
          /db_xref="taxon:3702"
FEATURES
Source

```



```

/chromosome="3"
/clone="U25336"
/note="This clone is in pENTR/SD-topo. This is a cloned
PCR product using RIKEN clone RAL09-83-B22 (AY099706) as
a template
ecotype: Columbia"
1..1092
/gene="At3g09150"
/note="F3L24.1"
1..990
/gene="At3g09150"
/codon_start=1
/product="unknown protein"
/protein_id="AA01300.1"
/db_xref="GI:22136444"
/translacion="MALSMERFSGCFRNPVILISASPKINTLRKRKRL
KSAVSTKREFALEETKRIVLEPSHLOKISSMTGLDLEQLQMAEFSKRL
RSMALNETMVFDFAGMEPEYDPIFCANFPTNVNIVLDLNLPHLDQTDYO
DKRYNKLMSIYHKAEYFPMGKLTGSIKFSPVMTFRSSSEKHKALFSALEY
YQAVLEMTIOVREEMERSHVANCEAOHKYLTWRADKDPGHGLKRLGKAKELLR
DFLENGVDELGTFFIDYFPEYOREDETVSDKRSIIKSYETREMDLTGPTG"
BASE COUNT      352 a      194 c      253 g      293 t
ORIGIN
Query Match      12.9%; Score 278; DB 8; Length 1092;
Best Local Similarity 100.0%; Pred. No. 9,1e-138;
Matches 278; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1741 AGAGCTGCTNAGGAGATTCTGTTCAATGGGGTGGATGAGTGAACAAACATTCA 1800
Db 815 AGGAGCTGCTNAGGAGATTCTGTTCAATGGGGTGGATGAGTGAACAAACATTCA 874
OY 1801 TTGATTTCTTCGAGAGACCAACAGAGATGAGTGAACGATTAACGAGTATCA 1860
Db 875 TTGATTTCTTCGAGAGACCAACAGAGATGAGTGAACGATTAACGAGTATCA 934
OY 1861 TTGGAGAGTCATATGAACCTGCTCCATGGATTTAACAGCAATTTATCGGCTAACAT 1920
Db 935 TTGGAGAGTCATATGAACCTGCTCCATGGATTTAACAGCAATTTATCGGCTAACAT 994
OY 1921 GATATATGTGAACAGTCATTTTCAGATCATCAACAGAGAGCTGAACCTTAGGGA 1980
Db 995 GATATATGTGAACAGTCATTTTCAGATCATCAACAGAGAGCTGAACCTTAGGGA 1054
OY 1981 AGTAGAATAAGAAAGAGCAGATGAGTCTCTCAG 2018
Db 1055 AGTAGAATAAGAAAGAGCAGATGAGTCTCTCAG 1092
RESULT 7
ATAC009326/c 85561 bp DNA linear PLN 24-JAN-2001
LOCUS Arabidopsis thaliana chromosome III P1 MZB10 genomic sequence,
ACCESSION AC009326
VERSION AC009326.8 GI:12408713
KEYWORDS HTG.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
REFERENCE 1 (bases 1 to 85561)
AUTHORS Loh, X., Kaul, S., Town, C.D., Benito, M.-I., Creasy, T.H., Haas, B.,
Rohring, C.M., Koo, H., Fujii, C.Y., Utterback, T.R., Barnstead, M.E.,
Bowman, C.L., White, O., Niernman, W.C. and Fraser, C.M.
TITLE Arabidopsis thaliana chromosome III P1 MZB10 genomic sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 85561)
AUTHORS Loh, X. and Kaul, S.
TITLE Direct Submission
JOURNAL Submitted (16-AUG-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlinetlgr.org

```

```

REFERENCE 3 (bases 1 to 85561)
AUTHORS Loh, X.
TITLE Direct Submission
JOURNAL Submitted (24-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
COMMENT On Jan 24, 2001 this sequence version replaced gi.12280842.
Address all correspondence to:
Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr.
Rockville, MD 20850, USA
e-mail: xlinetlgr.org
P1 clone MZB10 is from Arabidopsis chromosome III and is near the
molecular marker C1C7A12.
The orientation of the sequence is from SP6 to T7 end of the P1
clone.
Genes were identified by a combination of three methods: Gene
prediction programs including GRAIL (available by anonymous ftp
from arthur.gem.ornl.gov), Genefinder (Phil Green, University of
Washington), Genscan (Chris Burge
http://genome.stanford.edu/~chris/GENSCANW.html), and NetPlantene
(http://www.cds.dtu.dk/netpgene/obsnetpgene.html), searches of the
complete sequence against a peptide database and the Arabidopsis
EST database at TIGR (http://www.tigr.org/cdb/at.html).
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity are named as "unknown" proteins.
Genes without protein or EST similarity, that are predicted by more
than two gene prediction programs over most of their length are
annotated as "hypothetical" proteins. Genes encoding tRNAs are
predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE). Simple repeats are
identified by RepeatMasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
genomic sequence that are not annotated as genes but have predicted
exons by GRAIL are annotated as misc features.
location/Qualifiers
1..85561
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="III"
/map="C1C7A12"
/clone="MZB10"
349..702
/note="exon predicted by xgrail, quality excellent"
complement(587..681)
/note="exon predicted by xgrail, quality
excellent_shadowexon"
881..1171
/note="exon predicted by xgrail, quality good"
1508..1606
/note="exon predicted by xgrail, quality marginal"
complement(1516..1602)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
1653..1758
/note="exon predicted by xgrail, quality marginal"
complement(1682..1731)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
1854..1933
/note="exon predicted by xgrail, quality good_shadowexon"
complement(1852..1922)
/note="exon predicted by xgrail, quality good_shadowexon"
1933
/note="exon predicted by xgrail, quality marginal"
<2185..>3281
/gene="MZB10.1"
/note="similar to MITOCHONDRIAL INNER MEMBRANE PROTEASE
SUBUNIT 2 GB:P46972 (Saccharomyces cerevisiae)"
join(<2185..2338,2432..2504,2922..3081,3204..3281)
/gene="MZB10.1"
join(2185..2338,2432..2504,2922..3081,3204..3281)
CDS

```

```

misc_feature
      misc_feature
            misc_feature
                  gene
                        mRNA
                              CDS
                                      Query Match
                                          Best Local Similarity    1.28; Score 26; DB 8; Length 85561;
                                                Matches     26; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Ox       1550 TGCAATGTTTGACAGACACTGTATCT 1575
          |||||||TGGTTCAGCAGCACACTGTATCT 21688
          |||||||TGGTTCAGCAGCACACTGTATCT 21688

RESULT 8
GP1P451A1/c           2674 bp.  mRNA   linear  ROD 26-MAY-1999
LOCUS                 GP1P451A1
DEFINITION             Cavia porcellus mRNA for cytochrome P450 1A1, complete cds.
ACCESSION              D11043
VERSION                D11043.1 GI:220305
KEYWORDS               Cytochrome P450 1A1.
SOURCE                 Cavia porcellus (strain:Hartley) CDNA to mRNA, clone:GPC1.
ORGANISM               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                       Mammalia; Eutheria; Rodentia; Hystriocognathi; Caviidae; Cavia.
REFERENCE              Ohgura,S., Ishizaki,K. and Shniriki,N.
TITLE                  Molecular cloning of guinea pig cyp1a1: complete primary structure
AUTHORS                and fast mobility of expressed protein on electrophoresis
JOURNAL               Biochim. Biophys. Acta 1216 (2), 237-244 (1993)
MEDLINE               94060096
REFERENCE              2 (bases 1 to 2674)
AUTHORS                Ishizaki,K.
TITLE                  Direct Submission
JOURNAL               Submitted (24-Apr-1992) Kozo Ishizaki, Government Industrial
                       Development Laboratory-Hokkaido, Applied Chemistry; 2-17-2-1
                       Tsukisamu-Higashi Toyohira-Ku, Sapporo, Hokkaido 062, Japan
                       (Tel:011-857-8923, Fax:011-857-8900)
FEATURES
source                Location/Qualifiers
                      1..2674
                      /organism="Cavia porcellus"
                      /strain="Hartley"
                      /db_xref="taxon:10141"
                      /clone="GPC1"
                      126..1676

```

/codon_start=1
 /product="cytochrome P450 1A1"
 /protein_id="AA01794.1"
 /db_xref="GI:220306"
 /translation="MSTSAEMLLTATITFCIVLVWRIEPRQVKGKSPGPGWPL
 IGHMLTGTGNPHALITRLSAYGVDLQIRISTVVVLSGDIITRLQALVRGDDPKGR
 PDLSFTFISDQMTNPDSGPVMAARLAQSAKSFVSPADPVSQSLSEHVK
 KEAYLTKRFOELMAGQHPDPRVYVYVAVNSAICFGGRSHDDQQLLELDLN
 EFGVSSGNSDFTILIRYPSATMDPEFDLNRSEVFIQKMKREHYKTEKHIRD
 ITDSLIEHCODRKLDKNANIOISQKTIIGYLDLFGAGFDITTTAISLDTLVNPR
 IOKRIQELDLVIGEROPOLADRPKLYPMFAFISEVFRSSMPEFTIPHSTTDTSL
 NGFYIPRGCCIFVNMQIINHOKLMDGSPVAFRPRPLSPDGVKALSEKTIIGLGR
 RRCLEGVIGRWEVFLFALLIQLEFSTSPGVKIDMPYGLTWKYSRCEHFQACTRP
 FVLCPEA"

polya_signal 2656.2661
 BASE COUNT 640 a 686 c 661 g 687 t
 ORIGIN

Query Match 1.1%; Score 23; DB 10; Length 2674;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1812 CCAGAGTACCAACAGAGATG 1834
 ||||||||||||||||||
 DB 185 CCAGAGTACCAACAGAGATG 163

RESULT 9
 LOCUS S79865/c
 DEFINITION pig-A-synthetic element required to initiate glycolipid
 phospholipid (GPI) anchor assembly [exon 1, promoter] [mice,
 genomic, 1077 nt].
 ACCESSION S79865
 VERSION S79865.1 GI:1195529
 KEYWORDS
 SOURCE Mus sp.
 ORGANISM Mus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1077)
 AUTHORS Yu, J. and Medof, M.E.
 TITLE Characterization of the murine pig-A promoter region high
 constitutive pig-A gene expression in brain
 JOURNAL Blochem. Biophys. Res. Commun. 215 (2), 497-503 (1995)
 MEDLINE 96011806
 PUBMED 7487983
 REMARK GenBank staff at the National Library of Medicine created this
 entry [NCBI gidsq 172336] from the original journal article.
 This sequence comes from Fig. 2.

FEATURES
 source location/qualifiers
 1..1077
 /organism="Mus sp."
 /db_xref="taxon:10095"

gene 1..1077
 /partial
 /gene="pig-A"
 /note="synthetic element required to initiate
 glycolipid phospholipid (GPI) anchor assembly"
 938..>997
 mRNA /gene="pig-A"

BASE COUNT 226 a 288 c 296 g 267 t
 ORIGIN

Query Match 1.0%; Score 22; DB 10; Length 1077;
 Best Local Similarity 100.0%; Pred. No. 6.7;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0.;

OY 610 TAAAGTCAGATTGTTT 631
 ||||||||||||||||||
 DB 246 TAAAGTCAGATTGTTT 225

RESULT 10
 LOCUS AL663025
 DEFINITION Mouse DNA sequence from clone RP23-404B19 on chromosome 11,
 complete sequence.

ACCESSION AL663025
 VERSION AL663025.7 GI:19572091
 KEYWORDS HTG.

SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 101305)
 AUTHORS Tracey A.
 TITLE Direct Submission
 JOURNAL Submitted (14-MAR-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humberty@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On May 21, 2002 this sequence version replaced gi:18477424.

COMMENT

During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP
 database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-404B19 is
 from the RPI-23 Mouse PAC library
 constructed by the group of Pieter de Jong.
 For further details see http://www.chori.org/bacpac/home.htm
 VECTOR: pACe3.6

IMPORTANT: This sequence is not the entire insert of clone
 RP23-404B19 it may be shorter because we sequenced overlapping
 sections only once, except for a short overlap.

The true left end of clone RP23-404B19 is at 1 in this sequence.
 The true left end of clone RP23-387J5 is at 99306 in this sequence.
 The true right end of clone RP23-20A9 is at 59853 in this sequence.

FEATURES
 source location/qualifiers
 1..101305
 /organism="Mus musculus"

/db_xref="taxon:10090"
 /chromosome="11"
 /clone="RP23-404B19"
 /clone_11b="RPI-23"
 22091..22100
 /note="Sequence from uni-directional dGTP big dye
 terminator reads only"

misc_feature
 BASE COUNT 27202 a 23727 c 23324 g 27052 t
 ORIGIN

Query Match 1.0%; Score 22; DB 10; Length 101305;
 Best Local Similarity 100.0%; Pred. No. 5;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 351 TGAGAGGAGAAAGAAAGATT 372
 ||||||||||||||||||
 DB 100591 TGAGAGGAGAAAGAAAGATT 100612

RESULT 11
 LOCUS AL732475
 DEFINITION Mouse DNA sequence from clone RP23-202H24 on chromosome X, complete

sequence.
 ACCESSION AL732475 GI:21615718
 VERSION AL732475.6
 KEYWORDS HTG.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 165720)
 Phillimore, B.
 TITLE Direct Submission
 AUTHORS Submitted (25-JUN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 COMMENT On Jun 26, 2002 this sequence version replaced gi:21614706.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquerry@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPep; Information on the WormPep database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-202H24 is from the RPCR-23 Mouse PAC library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACe3.6.
 Location/Qualifiers
 1..165720
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="X"
 /clone_lib="RPCR-23"
 /clone_1lb="RPCR-23"
 BASE COUNT 46457 a 33761 c 34398 g 51104 t
 ORIGIN
 Query Match 1.0%; Score 22; DB 10; Length 165720;
 Best Local Similarity 100.0%; Pred. No. 4.8;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 610 TAAAGTCAGATTGTTT 631
 Db 109491 TAAAGTCAGATTGTTT 109470
 RESULT 12
 AL645626
 LOCUS AL645626 170821 bp DNA linear ROD 12-JUN-2002
 DEFINITION Mouse DNA sequence from clone RP23-397J5 on chromosome 11, complete
 sequence.
 ACCESSION AL645626
 VERSION AL645626.10 GI:21425601
 KEYWORDS HTG.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 170821)
 Whitehead, S.
 TITLE Direct Submission
 AUTHORS Submitted (12-JUN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 COMMENT On Jun 13, 2002 this sequence version replaced gi:19572018.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPep; Information on the WormPep database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-397J5 is from the RPCR-23 Mouse PAC library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACe3.6.
 Location/Qualifiers
 1..170821
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="11"
 /clone_lib="RPCR-23"
 /clone_1lb="RPCR-23"
 BASE COUNT 45754 a 39003 c 39852 g 46212 t
 ORIGIN
 Query Match 1.0%; Score 22; DB 10; Length 170821;
 Best Local Similarity 100.0%; Pred. No. 4.8;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 351 TGAGAGGAGAGAGAAAGATT 372
 Db 1286 TGAGAGGAGAGAGAAAGATT 1307
 RESULT 13
 AC115692/c
 LOCUS AC115692 176725 bp DNA linear HTG 26-JUN-2002
 DEFINITION Mus musculus clone RP23-463G11, WORKING DRAFT SEQUENCE, 11 ordered pieces.
 ACCESSION AC115692
 VERSION AC115692.2 GI:21591964
 KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE HTG; house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 176725)
 Birren, B., Nussbaum, C. and Lander, E.
 TITLE Unpublished
 AUTHORS Submitted (2 (bases 1 to 176725))
 REFERENCES
 Birren, B., Linton, L., Nussbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, D., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dekrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fairo, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Glinde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,

SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 221729)
AUTHORS Phillimore, B.
TITLE Direct Submission
JOURNAL Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humphreys@sanger.ac.uk
COMMENT On Aug 21, 2002 this sequence version replaced gi:22265580.

Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humphreys@sanger.ac.uk

----- Project Information -----
Center project name: DM13K9

----- Summary Statistics -----
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 9% of reads
Consensus quality: 221373 bases at least Q40
Consensus quality: 221407 bases at least Q30
Consensus quality: 221414 bases at least Q20
Insert size: 221429; sum-of-coverage
Insert size: 225092; 3.1% error; agarose-fp
Quality coverage: 7.94x in Q20 bases; sum-of-coverage
Coverage: 7.85x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 126679: contig of 126679 bp in length
* 126680 126779: gap of 100 bp
* 126780 144377: contig of 17598 bp in length
* 144378 144477: gap of 100 bp
* 144478 184550: contig of 40073 bp in length
* 184551 184650: gap of 100 bp
* 184651 221729: contig of 37079 bp in length.

location/Qualifiers
1. 221729
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-113K9"
/clone_lib="RPCI-23"
1. 126679
/note="assembly-fragment:02731
fragment_chain:1"
126780. 144377
/note="assembly-fragment:01671
fragment_chain:1"
144478. 184550
/note="assembly-fragment:03107
fragment_chain:1"
184651. 221729
/note="assembly-fragment:00406
fragment_chain:1"

BASE COUNT 65809 a 46520 c 46332 g 62768 t 300 others

ORIGIN

Query Match 1.0%; Score 22; DB 2; Length 221729;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

529 CTGTGAATTTGGGTAAATTGA 550
|||||

Db 95515 CTGTGAATTTGGGTAAATTGA 95494

RESULT 15
AC125049
LOCUS

DEFINITION Mus musculus chromosome UNK clone RP23-367C15, WORKING DRAFT
AC125049
SEQUENCE 14 unordered pieces.
AC125049 1 GI:21490490
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

ACCESSION
AC125049
VERSION
AC125049.1
KEYWORDS
SOURCE
ORGANISM
house mouse.

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@wustl.wustl.edu
Center project name: M.BA0367C15

----- Project Information -----
Center project name: M.BA0367C15

----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-terminator; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 223685 bases at least Q40
Consensus quality: 224340 bases at least Q30
Consensus quality: 224800 bases at least Q20
Insert size: 216000; agarose-fp
Insert size: 227586; sum-of-coverage
Quality coverage: 10.66 in Q20 bases; sum-of-coverage
Quality coverage: 10.66 in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1091: contig of 1091 bp in length
* 1092 1191: gap of unknown length
* 1192 2667: contig of 1476 bp in length
* 2668 2767: gap of unknown length
* 2768 4268: contig of 1501 bp in length
* 4269 4368: gap of unknown length
* 4369 5869: contig of 1501 bp in length
* 5870 5970: gap of unknown length
* 5971 7852: contig of 1883 bp in length
* 7853 7952: gap of unknown length
* 7953 10584: contig of 2632 bp in length
* 10585 10684: gap of unknown length
* 10685 15440: contig of 4756 bp in length
* 15441 15540: gap of unknown length
* 15541 18658: contig of 3118 bp in length
* 18659 18758: gap of unknown length
* 18759 25482: contig of 6724 bp in length
* 25483 25582: gap of unknown length

```
* 25583 37959: contig of 12377 bp in length
* 37960 38059: gap of unknown length
* 38060 62481: contig of 24422 bp in length
* 62482 62581: gap of unknown length
* 62582 9814: contig of 35533 bp in length
* 98115 98214: gap of unknown length
* 98215 227712: contig of 129498 bp in length
* 227713 227812: gap of unknown length
* 227813 228443: contig of 631 bp in length.
```

FEATURES

source

1. .228443

Location/Qualifiers

/organism="Mus musculus"

/db_xref="taxon:10090"

/chromosome="UNK"

/clone="RP23-367C15"

1. .1091

/note="assembly_name:Contig17"

1192. .2667

/note="assembly_name:Contig43"

2768. .4268

/note="assembly_name:Contig45"

4369. .5869

/note="assembly_name:Contig49"

5970. .7852

/note="assembly_name:Contig50"

7953. .10584

/note="assembly_name:Contig51"

10685. .15440

/note="assembly_name:Contig52"

15541. .18658

/note="assembly_name:Contig53"

18759. .25482

/note="assembly_name:Contig54"

23583. .37959

/note="assembly_name:Contig55"

38060. .62481

/note="assembly_name:Contig56"

62582. .98114

/note="assembly_name:Contig57"

98215. .227712

/note="assembly_name:Contig58"

227813. .228443

/note="assembly_name:Contig14"

BASE COUNT 66846 a 47680 c 46061 g 66469 t 1387 others

ORIGIN

Query Match 1.0%; Score 22; DB 2; Length 228443;

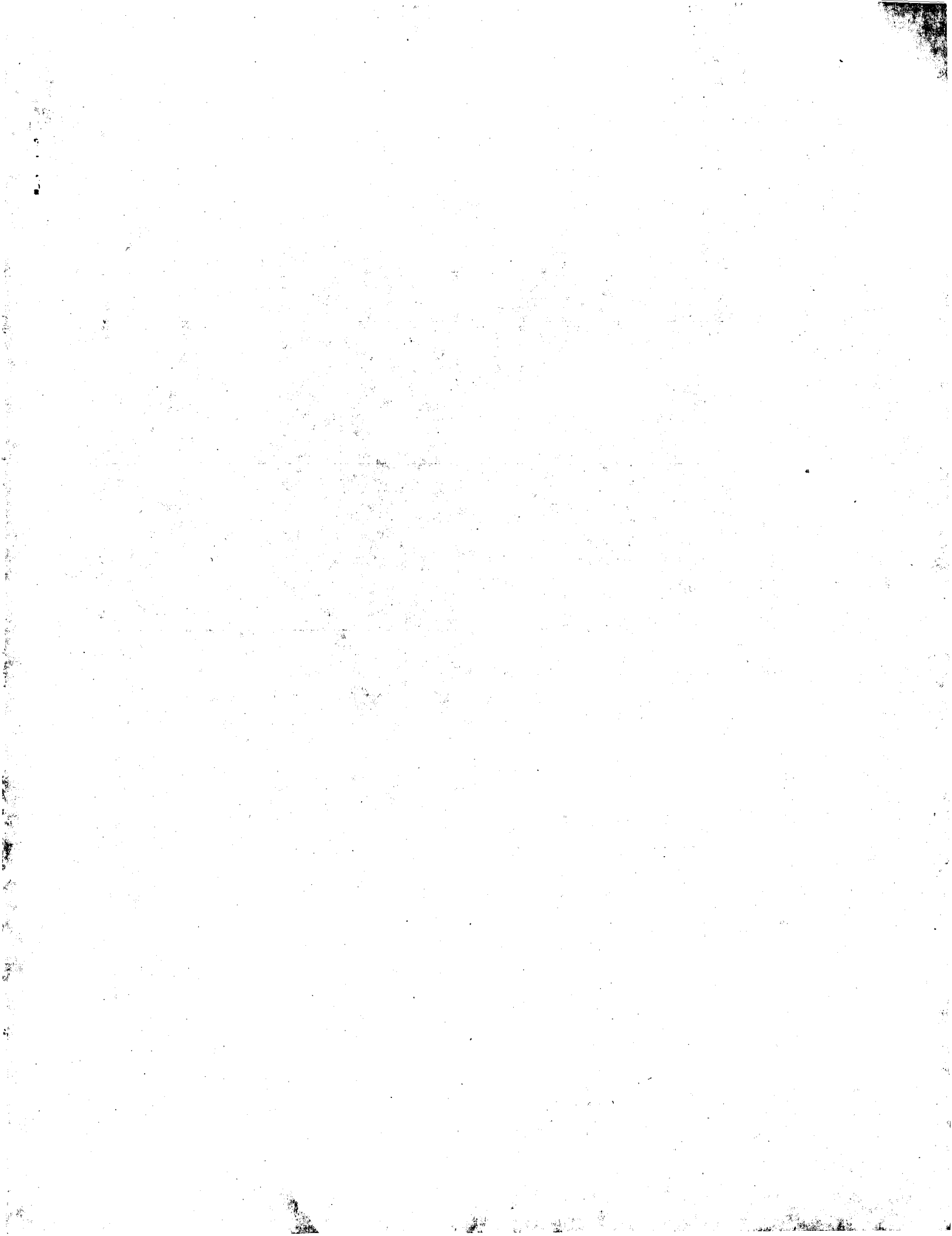
Best Local Similarity 100.0%; Pred.No. 4.7;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 351 TGAGAGGAGAGAGAGAGATT 372

DB 178518 TGAGAGGAGAGAGAGAGATT 178539

Search completed: June 24, 2003, 16:43:03
job time : 3717 secs



ET /numb

Arabidopsis Hy2 lo
Human secreted pro
Tumour suppressor
Signal transductio
Human ORF4403 cDNA
Human cancer relat
Genomic amplicon o
Human cancer relat
Human prostate exp
Human cancer relat
Arabidopsis thaila
Mouse type II coll
Envelope protein g
cMY RNA4 leader se
Arabidopsis thaila
Arabidopsis thaila
Tomato DHS coding
Tomato leaf deoxyh
Construct compri
Drosophila Virilis
Kidney cancer rela
Human angiogenesis
DNA transcription
Plasmid p20534 for
Plasmid p20534 for
Enterococcus faeca
Chemically treated
Human immune syste
Human nervous syst
Soybean J151002 reg
Nucleotide sequen
Methanococcus jan
Arabidopsis mature
Human foetal liver
Probe #17156 for g
Human brain expres
Human bone marrow

Accession	Position	Variant	Annotation
FT	770..851	/*tag= f	Intron
FT	/number= 2		
FT	852..946	/*tag= g	exon
FT	/number= 3		
FT	947..1032	/*tag= h	Intron
FT	/number= 3		
FT	1033..1126	/*tag= i	exon
FT	/number= 4		
FT	1127..1212	/*tag= j	Intron
FT	/number= 5		
FT	1213..1344	/*tag= k	exon
FT	/number= 5		
FT	1345..1418	/*tag= l	Intron
FT	/number= 5		
FT	1419..1523	/*tag= m	exon
FT	/number= 6		
FT	1524..1611	/*tag= n	Intron
FT	/number= 6		
FT	1612..1662	/*tag= o	exon
FT	/number= 7		
FT	1663..1742	/*tag= p	Intron
FT	/number= 7		
FT	1742..2148	/*tag= q	exon
FT	/number= 8		
FT	replace(234,Tt)		Variation
FT	/*tag= x		
FT	/standard_name= "Single nucleotide polymorphism"		
FT	/note= "inserted T at nucleotide 234 in Landsberg erecta and Massilewskija ecotypes"		
FT	replace(364,T)		Variation
FT	/*tag= s		
FT	/standard_name= "Single nucleotide polymorphism"		
FT	/note= "amino acid change to Asn in Landsberg erecta and Massilewskija ecotypes"		
FT	replace(515,A)		Variation
FT	/*tag= t		
FT	/standard_name= "Single nucleotide polymorphism"		
FT	/note= "in intron, Landsberg erecta ecotype"		
FT	replace(844,A)		Variation
FT	/*tag= u		
FT	/standard_name= "Single nucleotide polymorphism"		
FT	/note= "silent, Landsberg erecta ecotype"		
FT	replace(1145,T)		Variation
FT	/*tag= v		
FT	/standard_name= "Single nucleotide polymorphism"		
FT	/note= "in intron, Landsberg erecta ecotype"		
FT	replace(1910,T)		Variation
FT	/*tag= w		
FT	/standard_name= "Single nucleotide polymorphism"		
FT	/note= "silent, Massilewskija ecotype"		
PN	W0200194548-A2		
XX	13-DEC-2001		
XX	05-JUN-2001; 2001WO-US18326		
XX	08-JUN-2000; 2000US-210286P		
XX	26-FEB-2001; 2001US-271758P		
XX	29-MAY-2001; 2001US-0210286		

PA	(REGC) UNIV CALIFORNIA.
PI	Lagararias JC, Kochi T, Frankenberg N, Gambetta GA, Montgomery BL;
XX	
DR	WPI; 2002-195566/25.
XX	P-PsDB; AAM50863.
XX	
PT	Novel isolated HY2 family bilin reductase having bilin reductase
PT	activity, useful for converting biliverdin to phytylbilin, and for
PS	producing a photoactive holophytochrome and/or phytofluor
XX	
PS	Example 3; Fig 3B; 102pp; English.
XX	
CC	The present sequence is that of the HY2 gene of Arabidopsis thaliana
CC	ecotype Columbia on chromosome 3, with single nucleotide
CC	polymorphisms found in ecotypes Landsberg erecta and Wassilewskija
CC	indicated. A positional cloning strategy was used to isolate the
CC	HY2 gene. For fine mapping, an hyz-1 mutant of ecotype Landsberg
CC	erecta was outcrossed with wild-type ecotype Columbia, and a
CC	mapping population was selected from F2 families with a long
CC	hypocotyl phenotype. This localised from F2 genes to 2 overlapping
CC	bacterial artificial chromosome clones, and the gene was finally
CC	identified by sequencing of wild-type and mutant alleles. The HY2
CC	gene encodes a ferredoxin-dependent biliverdin reductase,
CC	phytylchromobilin synthase (see AAM50863), that is related to a
CC	family of proteins found in oxygenic photosynthetic bacteria. The
CC	HY2 gene product is localised in the chloroplast. HY2 is an
CC	example of bilin reductases of the invention, which are useful e.g.
CC	for the conversion of biliverdin to phytylbilin and the assembly of
CC	holophytochromes or phytoflavins. The cloning of the Arabidopsis
CC	HY2 gene will help to identify phytylchromobilin synthase genes from
CC	other plant species. A claimed method of producing a photoactive
CC	holophytochrome involves co-expressing a haem oxygenase, an
CC	apophytochrome and a ferredoxin-dependent bilin reductase in a
CC	cell, where the cell produces the photoactive holophytochrome and
CC	where the apophytochrome and/or the bilin reductase are expressed
CC	by heterologous nucleic acids. The cell may be an algal, yeast,
CC	bacterial, plant, insect or mammalian cell, and the bilin reductase
CC	is preferably an HY2 family bilin reductase.
XX	
SQ	Sequence 2160 BP; 645 A; 361 C; 461 G; 693 T; 0 other;
Query Match	100.0%; Score 2160; DB 24; Length 2160;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 2160; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 GAATTCGCCACGCTCAAGCTGACTGTGCATTCCACGTGGCGATGGGCCCTTATAGTTGG 60
DB	1 GAATTCGCCACGCTCAAGCTGACTGTGCATTCCACGTGGCGATGGGCCCTTATAGTTGG 60
OY	61 ACCATGACTCGGACGCGANGTTGAATTCATTGTGCTGCCAATGCGTTTGTCACATGA 120
DB	61 ACCATGACTCGGACGCGANGTTGAATTCATTGTGCTGCCAATGCGTTTGTCACATGA 120
OY	121 AACGTGTAAATTTTATCTCTTTTATACATAAAGAATCTGCTTTTTCACATTTTCAGTA 180
DB	121 AACGTGTAAATTTTATCTCTTTTATACATAAAGAATCTGCTTTTTCACATTTTCAGTA 180
OY	181 TGAAGAAGAAATTTGAAGAGAGTGTCCGAGGAAGAGACCTTTGGTTTCAGTTTGGAGCT 240
DB	181 TGAAGAAGAAATTTGAAGAGAGTGTCCGAGGAAGAGACCTTTGGTTTCAGTTTGGAGCT 240
OY	241 TGTTGTATATGCTTTATCATAGGAGTTGGGTTTTCAATTTGGTTCATGGGTTCAGGACACC 300
DB	241 TGTTGTATATGCTTTATCATAGGAGTTGGGTTTTCAATTTGGTTCATGGGTTCAGGACACC 300
OY	301 AAACCCACCTGTTTAACTCTCTGCAACCCCAATTAATCAATTCACGTTGAGAAGAG 360
DB	301 AAACCCACCTGTTTAACTCTCTGCAACCCCAATTAATCAATTCACGTTGAGAAGAG 360
OY	361 AAAAGAAAAGATCTTACTTAGAGTCTGCTGTGCTGTATTAAGAAATTCGACAGTCTGC 420
DB	361 AAAAGAAAAGATCTTACTTAGAGTCTGCTGTGCTGTATTAAGAAATTCGACAGTCTGC 420

QY 421 TTTAGAGAAACGAGAAAGATCGTTCGAACTTCACATCTCCAGTATATGCAAT 480
DB 421 TTTAGAGAAACGAGAAAGATCGTTCGAACTTCACATCTCCAGTATATGCAAT 480
QY 481 TACATTTGCTGATAGTGTGAGTGGAGATTAATTTCTCATGTTTCTTGCCTGTAATTTG 540
DB 481 TACATTTGCTGATAGTGTGAGTGGAGATTAATTTCTCATGTTTCTTGCCTGTAATTTG 540
QY 541 GGTAAATGATTTAGTGTGATAGGACCAACCAATTAATCTTACTGTTATAGACTGC 600
DB 541 GGTAAATGATTTAGTGTGATAGGACCAACCAATTAATCTTACTGTTATAGACTGC 600
QY 601 TTATATAGTAAAGTTCAGATTTGTTTCTTCAATCAGAACTGTTTCAGAAAGTA 660
DB 601 TTATATAGTAAAGTTCAGATTTGTTTCTTCAATCAGAACTGTTTCAGAAAGTA 660
QY 661 TAGTAGCATGACAGGACTAGATGTAAGACCGAACTTCAATGCTTGTAAATCTTC 720
DB 661 TAGTAGCATGACAGGACTAGATGTAAGACCGAACTTCAATGCTTGTAAATCTTC 720
QY 721 AAAATATGACTCTTGAGAGATATGSCAATAGACAATGACAACTGTTAACTTCA 780
DB 721 AAAATATGACTCTTGAGAGATATGSCAATAGACAATGACAACTGTTAACTTCA 780
QY 781 GCAGTACAAACTGATGCTTGTAGTCCCAATTCCTTACTTCAATGATGATGTTGTA 840
DB 781 GCAGTACAAACTGATGCTTGTAGTCCCAATTCCTTACTTCAATGATGATGTTGTA 840
QY 841 TCTTCGCTTAGGCTTTGACTTTGCGGGTTTCATGAGACCTGAGTATGATCTCCAT 900
DB 841 TCTTCGCTTAGGCTTTGACTTTGCGGGTTTCATGAGACCTGAGTATGATCTCCAT 900
QY 901 TCTGCTGAATCTTTTCAATCTTCAACAGCTTAACATAGTGTATGTAATCTTCT 960
DB 901 TCTGCTGAATCTTTTCAATCTTCAACAGCTTAACATAGTGTATGTAATCTTCT 960
QY 961 AGTATGCTGAGTATATCAGGCTGTATGTCGAACAGTATGATGATGATGATGAT 1020
DB 961 AGTATGCTGAGTATATCAGGCTGTATGTCGAACAGTATGATGATGATGATGAT 1020
QY 1021 GTTCCTCTTAGGAGACCTTAATCTTGTGATCAGTGTGACGACGAGATTTACAGA 1080
DB 1021 GTTCCTCTTAGGAGACCTTAATCTTGTGATCAGTGTGACGAGATTTACAGA 1080
QY 1081 CAAGTATTTATACAGAAATATGTCATATATACAAATATCTGAGTGACACAAAGAT 1140
DB 1081 CAAGTATTTATACAGAAATATGTCATATATACAAATATCTGAGTGACACAAAGAT 1140
QY 1141 ACACCAATTTACTCAATTTGCAAGTAAACCTTAATGCTGAGTGTAAATGATGATCTGAG 1200
DB 1141 ACACCAATTTACTCAATTTGCAAGTAAACCTTAATGCTGAGTGTAAATGATGATCTGAG 1200
QY 1201 ATTTATTTGACACTTTCCATGAGGAGGAAATTTGATGTAATCAATGATTTTTC 1260
DB 1201 ATTTATTTGACACTTTCCATGAGGAGGAAATTTGATGTAATCAATGATTTTTC 1260
QY 1261 TCGGCTTTGAGTGTGAGTGTGATGTTTCTGCTGAGCAAAAGCAATTAAGGCTTTGTC 1320
DB 1261 TCGGCTTTGAGTGTGAGTGTGATGTTTCTGCTGAGCAAAAGCAATTAAGGCTTTGTC 1320
QY 1321 TCTGCTTTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
DB 1321 TCTGCTTTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
QY 1381 GAAACTTTGACTGTGAATCTATCTATCTTCTTCAACAGGCTGAGATGACAAATCC 1440
DB 1381 GAAACTTTGACTGTGAATCTATCTATCTTCTTCAACAGGCTGAGATGACAAATCC 1440
QY 1441 AAGTAGAGGAGAGATGAGAACCATCTCATGTGAGAGGCAATTTGGAAGCACAAAGT 1500
DB 1441 AAGTAGAGGAGAGATGAGAACCATCTCATGTGAGAGGCAATTTGGAAGCACAAAGT 1500

QY 1501 ACTGACATGCGAGACACAAAGGTGATTCATTTCCCTTTTGTGTAATTTGCATGTTGA 1560
DB 1501 ACTGACATGCGAGACACAAAGGTGATTCATTTCCCTTTTGTGTAATTTGCATGTTGA 1560
QY 1561 ACAGACTGTATCTGATTTGTACATGATGATGATGATGATGATGATGATGATGATGAT 1620
DB 1561 ACAGACTGTATCTGATTTGTACATGATGATGATGATGATGATGATGATGATGATGAT 1620
QY 1621 CATGCTCTCTTAAAGTTAGTGTGACCAAGCAAGGATTAATTAAGATTTGATTC 1680
DB 1621 CATGCTCTCTTAAAGTTAGTGTGACCAAGCAAGGATTAATTAAGATTTGATTC 1680
QY 1681 CATGAGTGTCCCATTTATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
DB 1681 CATGAGTGTCCCATTTATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
QY 1741 AGGAGTGTCTAAGGATTTCTGTTCAATGGGGTGTAGTGTAGGACCAAAACATTTCA 1800
DB 1741 AGGAGTGTCTAAGGATTTCTGTTCAATGGGGTGTAGTGTAGGACCAAAACATTTCA 1800
QY 1801 TTGATTTACTTCCAGAGTACCAACAGAAAGTGTGAAGTGTGAAGTGTGAAGTGTGA 1860
DB 1801 TTGATTTACTTCCAGAGTACCAACAGAAAGTGTGAAGTGTGAAGTGTGAAGTGTGA 1860
QY 1861 TTGGGAAGTCAATATGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1920
DB 1861 TTGGGAAGTCAATATGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1920
QY 1921 GATATATGTGAACAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980
DB 1921 GATATATGTGAACAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980
QY 1981 AGTAGGAATTAAGAAAGGACAGTGTGATGATGATGATGATGATGATGATGATGATGAT 2040
DB 1981 AGTAGGAATTAAGAAAGGACAGTGTGATGATGATGATGATGATGATGATGATGATGAT 2040
QY 2041 TTGTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2100
DB 2041 TTGTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2100
QY 2101 GAGAACTCTGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160
DB 2101 GAGAACTCTGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160
RESULT 2
AAC38032
ID AAC38032 standard; DNA; 1308 BP.
XX AAC38032;
AC AAC38032;
DE 17-OCT-2000 (first entry)
XX
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 19544.
DE
XX Hybridisation assay: genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
PN
PN EP1033405-A2.
PD 06-SEP-2000.
PD
PE 25-FEB-2000; 2000EP-0301439.
PE
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123160.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-01300449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-01314449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 14-JUN-1999; 99US-0138540.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139617.
PR 23-JUN-1999; 99US-0139689.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 26-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142855.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144682.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 23-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.

```

PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 14.7%; Score 318; DB 21; Length 1308;

Best Local Similarity 99.5%; Pred. No. 1.4e-145;

Matches 418; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1741 AGGACCTGTAAGGATTTCTGTCATGAGGAGTGTAGGACACAAAACATTCA 1800
DB 878 AGGACCTGTAAGGATTTCTGTCATGAGGAGTGTAGGACACAAAACATTCA 937
QY 1801 TTGATTACTTCCAGATGACAAAGAGATGAGTGTAGGACATTAAGAGATATCA 1860
DB 938 TTGATTACTTCCAGATGACAAAGAGATGAGTGTAGGACATTAAGAGATATCA 997
QY 1861 TTGGAGATGATATGAACTGCTCCATGGATTTAAAGAGACATTTATCGCTAACAT 1920
DB 998 TTGGAGATGATATGAACTGCTCCATGGATTTAAAGAGACATTTATCGCTAACAT 1057
QY 1921 GATATATGTGAACAAGTCAAGTTCAGATCATCAACACAGAGACGTAAGGGA 1980
DB 1058 GATATATGTGAACAAGTCAAGTTCAGATCATCAACACAGAGACGTAAGGGA 1117
QY 1981 AGTAGAATTAAGAGACAGCAGATGAGAGTCTCAGGCTATGCTGATTTCAAGATGA 2040
DB 1118 AGTAGAATTAAGAGAGACAGCAGATGAGAGTCTCAGGCTATGCTGATTTCAAGATGA 1177
QY 2041 TTGTTGAGTACATGATGATGATTTTACAAAGTATGCTCAGCCCTTCATCAAAAT 2100
DB 1178 TTGTTGAGTACATGATGATGATTTTACAAAGTATGCTCAGCCCTTCATCAAAAT 1237
QY 2101 GAGATTCCTCGAGTATGATATGATTTTAATGAATGTTGCTGCTACCTAATCACA 2160
DB 1238 GAGATTCCTCGAGTATGATATGATTTTAATGAATGTTGCTGCTACCTAATCACA 1297

```

RESULT 3

ABA91752/c
ID ABA91752 standard; DNA; 36 BP.

AC ABA91752;

DT 07-MAY-2002 (first entry)

DE Arabidopsis blliverdin reductase HY2 gene PCR primer HY2Smalrev.

KW HY2; blliverdin reductase; phytochromobilin synthase; plant;

KW enzyme; PCR; primer; ss.

XX Arabidopsis thaliana.

OS WO200194548-A2.

XX 13-DEC-2001.

PF 05-JUN-2001; 2001WO-US18326.

XX 08-JUN-2000; 2000US-210286P.

PR 26-FEB-2001; 2001US-271758P.

PR 29-MAY-2001; 2001US-0210286.

PA (REGC) UNIV CALIFORNIA.

XX Lagarias JC, Kochl T, Frankenberger N, Gambetta GA, Montgomery BL;

XX WPI; 2002-195566/25.

XX Novel isolated HY2 family bilin reductase having bilin reductase

XX activity, useful for converting blliverdin to phytyobilin, and for

XX producing a photoactive holophytochrome and/or phytofluor

XX Example 1; Page 50; 102pp; English.

CC The present sequence is that of primer HY2Smalrev, which was used

CC with primer mHY2BglIIVd (see ABA91751) in the PCR amplification of

CC mHY2, the mature portion (lacking the predicted chloroplast transit

CC peptide sequence) of the HY2 gene (see ABA91766) of Arabidopsis

CC thaliana. The reverse primer contains a 5' SmaI site. The PCR

CC product was subcloned into the Escherichia coli expression vector

CC pGEX-6-P1 to give pGEX-mHY2. Recombinant HYB2 mature protein was

CC produced as a glutathione-S-transferase fusion, and purified by

CC affinity chromatography. The Arabidopsis HY2 gene encodes a

CC ferredoxin-dependent blliverdin reductase, phytochromobilin synthase

CC (see ABA90863), that is related to a family of proteins found in

CC oxygenic photosynthetic bacteria. HY2 is an example of HY bilin

CC reductases of the invention, which are useful e.g. for the

CC conversion of blliverdin to phytyobilin and the assembly of

CC holophytochromes or phytofluors.

CC Sequence 36 BP; 8 A; 10 C; 7 G; 11 T; 0 other;

CC

CC

CC

CC

CC

Query Match 1.3%; Score 28; DB 24; Length 36;

Best Local Similarity 100.0%; Pred. No. 0.0031;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1890 GATTTACAGACAAATTATCGGCTAAC 1917

DB 36 GATTTACAGACAAATTATCGGCTAAC 9

RESULT 4

ABA91743

ID ABA91743 standard; DNA; 22 BP.

AC ABA91743;

DT 07-MAY-2002 (first entry)

DE Arabidopsis chromosome 3 CAPS marker CMZB10.18 (HY2) PCR primer.

KW HY2; blliverdin reductase; phytochromobilin synthase; CAPS;

KW cleaved amplified polymorphic sequence; marker; plant; enzyme;

OS Arabidopsis thaliana.

XX WO200194548-A2.

XX 13-DEC-2001.

PF 05-JUN-2001; 2001MO-US18326.
 XX
 XX 08-JUN-2000; 2000US-210286P.
 PR 26-FEB-2001; 2001US-271758P.
 PR 29-MAY-2001; 2001US-0210286.
 XX
 PA (RECC) UNIV CALIFORNIA.
 XX
 PI Lagarias JC, Kochi T, Frankenberg N, Gambetta GA, Montgomery BL,
 XX
 DR WPI; 2002-195566/25.
 XX
 PT Novel isolated HY2 family bilin reductase having bilin reductase
 PT activity, useful for converting biliverdin to phytylobilin, and for
 PT producing a photoactive holophytochrome and/or phytofluor
 XX
 PS Example 1; Page 49; 102pp; English.
 XX
 CC The present sequence is that of a primer that was used, with the
 CC primer given in ABA91744, in the PCR amplification of the cleaved
 CC amplified polymorphic sequence (CAPS) marker CMZB10.18 of chromosome
 CC 3 of *Arabidopsis thaliana*. The primer pair includes a *DdeI*
 CC restriction endonuclease site. An *hy2-1* mutant of ecotype
 CC Landsberg erecta was outcrossed with wild-type ecotype Columbia,
 CC and a mapping population was selected from F2 families with a long
 CC hypocotyl phenotype. PCR primer pairs (see ABA91735-48) for 7 CAPS
 CC markers were used in a map-based cloning of the *HY2* gene. The *HY2*
 CC locus was initially mapped to an interval of about 66 kb between the
 CC markers CMZB10 and CFJ324. Fine mapping localised the *HY2* gene (see
 CC ABA91766) to 2 overlapping bacterial artificial chromosome clones,
 CC MZB10.18 and F3124.1. The *HY2* gene encodes a ferredoxin-dependent
 CC biliverdin reductase, phytychromobilin synthase (see AAM50863), that
 CC is related to a family of proteins found in oxygenic photosynthetic
 CC bacteria. *HY2* is an example of *HY* bilin reductases of the
 CC invention, which are useful e.g. for the conversion of biliverdin
 CC to phytylobilin and the assembly of holophytochromes or phytofluors.
 XX
 SO Sequence 22 BP; 7 A; 5 C; 4 G; 6 T; 0 other;
 Query Match 1.0%; Score 22; DB 24; Length 22;
 Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 762 CAATGCAGGTTTAACTTCAGCA 783
 DB 1 CAATGCAGGTTTAACTTCAGCA 22
 RESULT 5
 ID ABA91751 standard; DNA; 30 BP.
 AC ABA91751;
 XX
 DT 07-MAY-2002 (first entry)
 DE *Arabidopsis biliverdin reductase HY2 gene PCR primer mHY2Bg111fwd*.
 XX
 KW *HY2*; biliverdin reductase; phytychromobilin synthase; plant;
 KW enzyme; PCR; primer; ss.
 XX
 OS *Arabidopsis thaliana*.
 XX
 PN WO200194548-A2.
 PD 13-DEC-2001.
 XX
 PF 05-JUN-2001; 2001MO-US18326.
 XX
 PR 08-JUN-2000; 2000US-210286P.
 PR 26-FEB-2001; 2001US-271758P.
 PR 29-MAY-2001; 2001US-0210286.
 XX

PA (RECC) UNIV CALIFORNIA.
 XX
 PI Lagarias JC, Kochi T, Frankenberg N, Gambetta GA, Montgomery BL,
 XX
 DR WPI; 2002-195566/25.
 XX
 PT Novel isolated HY2 family bilin reductase having bilin reductase
 PT activity, useful for converting biliverdin to phytylobilin, and for
 PT producing a photoactive holophytochrome and/or phytofluor
 XX
 PS Example 1; Page 50; 102pp; English.
 XX
 CC The present sequence is that of primer mHY2Bg111fwd, which was used
 CC with primer HY2smarev (see ABA91752) in the PCR amplification of
 CC mHY2, the mature portion (lacking the predicted chloroplast transit
 CC peptide sequence) of the *HY2* gene (see ABA91766) of *Arabidopsis*
 CC *thaliana*. The forward primer contains a 5' *Bgl*II site. The PCR
 CC product was subcloned into the *Escherichia coli* expression vector
 CC pGEX-6-p1 to give pGEX-mHY2. Recombinant HY2 mature protein was
 CC produced as a glutathione-S-transferase fusion, and purified by
 CC affinity chromatography. The *Arabidopsis HY2* gene encodes a
 CC ferredoxin-dependent biliverdin reductase, phytychromobilin synthase
 CC (see AAM50863), that is related to a family of proteins found in
 CC oxygenic photosynthetic bacteria. *HY2* is an example of *HY* bilin
 CC reductases of the invention, which are useful e.g. for the
 CC conversion of biliverdin to phytylobilin and the assembly of
 CC holophytochromes or phytofluors.
 XX
 SO Sequence 30 BP; 6 A; 5 C; 9 G; 10 T; 0 other;
 Query Match 1.0%; Score 22; DB 24; Length 30;
 Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 383 GTCTGCTGTGTCGTATAGG 404
 DB 9 GTCTGCTGTGTCGTATAGG 30
 RESULT 6
 ID AAC42595 standard; DNA; 1232 BP.
 AC AAC42595;
 XX
 DT 17-OCT-2000 (first entry)
 DE *Arabidopsis thaliana* DNA fragment SEQ ID NO: 36142.
 XX
 KW Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.
 XX
 OS *Arabidopsis thaliana*.
 XX
 PN EP1033405-A2.
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0128845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999;	99US-0130510.
PR 23-APR-1999;	99US-0130891.
PR 28-APR-1999;	99US-0131448.
PR 30-APR-1999;	99US-0132040.
PR 30-APR-1999;	99US-0132407.
PR 04-MAY-1999;	99US-0132484.
PR 05-MAY-1999;	99US-0132485.
PR 06-MAY-1999;	99US-0132486.
PR 07-MAY-1999;	99US-0132487.
PR 11-MAY-1999;	99US-0132863.
PR 14-MAY-1999;	99US-0134256.
PR 14-MAY-1999;	99US-0134219.
PR 14-MAY-1999;	99US-0134221.
PR 14-MAY-1999;	99US-0134370.
PR 18-MAY-1999;	99US-0134768.
PR 19-MAY-1999;	99US-0134941.
PR 20-MAY-1999;	99US-0135124.
PR 21-MAY-1999;	99US-0135353.
PR 24-MAY-1999;	99US-0135629.
PR 25-MAY-1999;	99US-0136021.
PR 27-MAY-1999;	99US-0136392.
PR 28-MAY-1999;	99US-0136782.
PR 01-JUN-1999;	99US-0137222.
PR 03-JUN-1999;	99US-0137528.
PR 04-JUN-1999;	99US-0137502.
PR 07-JUN-1999;	99US-0137722.
PR 08-JUN-1999;	99US-0138094.
PR 10-JUN-1999;	99US-0138540.
PR 10-JUN-1999;	99US-0138847.
PR 14-JUN-1999;	99US-0139119.
PR 16-JUN-1999;	99US-0139452.
PR 16-JUN-1999;	99US-0139453.
PR 17-JUN-1999;	99US-0139492.
PR 18-JUN-1999;	99US-0139454.
PR 18-JUN-1999;	99US-0139455.
PR 18-JUN-1999;	99US-0139456.
PR 18-JUN-1999;	99US-0139457.
PR 18-JUN-1999;	99US-0139458.
PR 18-JUN-1999;	99US-0139459.
PR 18-JUN-1999;	99US-0139460.
PR 18-JUN-1999;	99US-0139461.
PR 18-JUN-1999;	99US-0139462.
PR 18-JUN-1999;	99US-0139463.
PR 18-JUN-1999;	99US-0139750.
PR 18-JUN-1999;	99US-0139763.
PR 21-JUN-1999;	99US-0139817.
PR 22-JUN-1999;	99US-0139839.
PR 23-JUN-1999;	99US-0140353.
PR 23-JUN-1999;	99US-0140354.
PR 24-JUN-1999;	99US-0140695.
PR 28-JUN-1999;	99US-0140823.
PR 29-JUN-1999;	99US-0140991.
PR 30-JUN-1999;	99US-0141287.
PR 01-JUL-1999;	99US-0141842.
PR 01-JUL-1999;	99US-0142154.
PR 02-JUL-1999;	99US-0142055.
PR 06-JUL-1999;	99US-0142390.
PR 08-JUL-1999;	99US-0142803.
PR 09-JUL-1999;	99US-0142920.
PR 12-JUL-1999;	99US-0142977.
PR 13-JUL-1999;	99US-0143542.
PR 14-JUL-1999;	99US-0143624.
PR 15-JUL-1999;	99US-0144005.
PR 16-JUL-1999;	99US-0144085.
PR 16-JUL-1999;	99US-0144086.
PR 19-JUL-1999;	99US-0144325.
PR 19-JUL-1999;	99US-0144331.
PR 19-JUL-1999;	99US-0144332.
PR 19-JUL-1999;	99US-0144333.
PR 19-JUL-1999;	99US-0144334.
PR 19-JUL-1999;	99US-0144335.
PR 20-JUL-1999;	99US-0144352.
PR 20-JUL-1999;	99US-0144632.
PR 20-JUL-1999;	99US-0144684.
PR 21-JUL-1999;	99US-0144814.
PR 21-JUL-1999;	99US-0145086.
PR 22-JUL-1999;	99US-0145085.
PR 22-JUL-1999;	99US-0145087.
PR 22-JUL-1999;	99US-0145089.
PR 22-JUL-1999;	99US-0145192.
PR 23-JUL-1999;	99US-0145145.
PR 23-JUL-1999;	99US-0145218.
PR 26-JUL-1999;	99US-0145224.
PR 27-JUL-1999;	99US-0145276.
PR 27-JUL-1999;	99US-0145913.
PR 27-JUL-1999;	99US-0145918.
PR 28-JUL-1999;	99US-0145919.
PR 02-AUG-1999;	99US-0145951.
PR 02-AUG-1999;	99US-0146386.
PR 02-AUG-1999;	99US-0146388.
PR 03-AUG-1999;	99US-0146389.
PR 04-AUG-1999;	99US-0147038.
PR 04-AUG-1999;	99US-0147204.
PR 05-AUG-1999;	99US-0147302.
PR 05-AUG-1999;	99US-0147192.
PR 06-AUG-1999;	99US-0147260.
PR 06-AUG-1999;	99US-0147303.
PR 09-AUG-1999;	99US-0147416.
PR 09-AUG-1999;	99US-0147493.
PR 10-AUG-1999;	99US-0147935.
PR 11-AUG-1999;	99US-0148319.
PR 12-AUG-1999;	99US-0148341.
PR 13-AUG-1999;	99US-0148565.
PR 13-AUG-1999;	99US-0148684.
PR 16-AUG-1999;	99US-0149368.
PR 17-AUG-1999;	99US-0149175.
PR 18-AUG-1999;	99US-0149426.
PR 20-AUG-1999;	99US-0149722.
PR 20-AUG-1999;	99US-0149723.
PR 20-AUG-1999;	99US-0149929.
PR 23-AUG-1999;	99US-0149930.
PR 23-AUG-1999;	99US-0149930.
PR 25-AUG-1999;	99US-0150566.
PR 26-AUG-1999;	99US-0150884.
PR 27-AUG-1999;	99US-0151065.
PR 27-AUG-1999;	99US-0151066.
PR 27-AUG-1999;	99US-0151080.
PR 30-AUG-1999;	99US-0151303.
PR 31-AUG-1999;	99US-0151438.
PR 01-SEP-1999;	99US-0151930.
PR 07-SEP-1999;	99US-0153363.
PR 10-SEP-1999;	99US-0153070.
PR 13-SEP-1999;	99US-0153758.
PR 15-SEP-1999;	99US-0154018.
PR 16-SEP-1999;	99US-0154029.
PR 20-SEP-1999;	99US-0154779.
PR 22-SEP-1999;	99US-0155133.
PR 23-SEP-1999;	99US-0155486.
PR 24-SEP-1999;	99US-0155659.
PR 28-SEP-1999;	99US-0156458.
PR 29-SEP-1999;	99US-0156596.
PR 04-OCT-1999;	99US-0157117.
PR 05-OCT-1999;	99US-0157753.
PR 06-OCT-1999;	99US-0157865.
PR 07-OCT-1999;	99US-0158029.
PR 08-OCT-1999;	99US-0158232.
PR 12-OCT-1999;	99US-0158369.
PR 13-OCT-1999;	99US-0158293.
PR 13-OCT-1999;	99US-0159295.
PR 13-OCT-1999;	99US-0159329.
PR 14-OCT-1999;	99US-0159330.
PR 14-OCT-1999;	99US-0159331.

PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160982.
PR 22-OCT-1999; 99US-0160983.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.0%; Score 21; DB 21; Length 1232;
Best Local Similarity 100.0%; Pred. No. 7.9;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1445 GAGGAGAGATGACCATC 1465
1140 GAGGAGAGATGACCATC 1160

Db

RESULT 7
AAC50780
AAC50780 standard; DNA; 1272 BP.

AC AAC50780;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 66110.

KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

PN EPI033405-A2.

PD 06-SEP-2000.

PE 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 23-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 28-APR-1999; 99US-0130891.

XX 30-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 04-MAY-1999; 99US-0132407.

XX 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134376.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142877.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.


```

PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160761.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

```

```

PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.0%; Score 21; DB 21; Length 1272;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1445 GAGGAGGAGATGACCATC 1465
DB 1180 GAGGAGGAGATGACCATC 1200

RESULT 8
ABA91744/c
ID ABA91744 standard; DNA; 20 BP.
XX
AC ABA91744;
XX
DT 07-MAY-2002 (first entry)
XX
DE Arabidopsis chromosome 3 CAPS marker CM2B10.18 (HY2) PCR primer.
XX
DE HY2; billyerdin reductase; phytochromobilin synthase; CAPS;
KW cleaved amplified polymorphic sequence; marker; plant; enzyme;
KM PCR; primer; ss.
XX
OS Arabidopsis thaliana.
XX
PN W0200194548-A2.
XX
PD 13-DEC-2001.
XX
PF 05-JUN-2001; 2001WO-US18326.
XX
PR 08-JUN-2000; 2000US-210286P.
PR 26-FEB-2001; 2001US-271758P.
PR 29-MAY-2001; 2001US-0210286.
XX
PA (REGC ) UNIV CALIFORNIA.
XX
PI Lagarias JC, Kochi T, Frankenberg N, Gambetta GA, Montgomery BJ,
XX WPI: 2002-195566/25.
XX
DR WPI: 2002-195566/25.
XX
PT Novel isolated HY2 family bilin reductase having bilin reductase
XX activity, useful for converting billyerdin to phytyobilin, and for
XX producing a photoactive holophytochrome and/or phytyofluor
XX
PS Example 1; Page 49; 102pp; English.
XX
XX
XX The present sequence is that of a primer that was used, with the
XX primer given in ABA91743, in the PCR amplification of the cleaved
XX amplified polymorphic sequence (CAPS) marker CM2B10.18 of chromosome
XX 3 of Arabidopsis thaliana. The primer pair includes a Ddel
XX restriction endonuclease site. An hy2-1 mutant of ecotype
XX Landsberg erecta was outcrossed with wild-type ecotype Columbia,
XX and a mapping population was selected from F2 families with a long
XX hypocotyl phenotype. PCR primer pairs (see ABA91735-48) for 7 CAPS
XX markers were used in a map-based cloning of the HY2 gene. The HY2
XX locus was initially mapped to an interval of about 66 kb between the

```

CC markers CM2B10 and cF3124. Fine mapping localised the HY2 gene (see
 CC ABA91766) to 2 overlapping bacterial artificial chromosome clones,
 CC MZB10.18 and F3124.1. The HY2 gene encodes a ferredoxin-dependent
 CC biliverdin reductase, phytychromobilin synthase (see AAM50863), that
 CC is related to a family of proteins found in oxygenic photosynthetic
 CC bacteria. HY2 is an example of HY bilin reductases of the
 CC invention, which are useful e.g. for the conversion of biliverdin
 CC to phytyobilin and the assembly of holophytochromes or phytyofluors.
 XX

SO Sequence 20 BP; 7 A; 4 C; 5 G; 4 T; 0 other;

Query Match 0.9%; Score 20; DB 24; Length 20;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1205 ATTGGAGACTTCCCATGG 1224
 Db 20 ATTGGAGACTTCCCATGG 1

RESULT 9

ABA91749
 ID ABA91749 standard; DNA; 20 BP.

AC ABA91749;

DT 07-MAY-2002 (first entry)

DE Arabidopsis HY2 locus PCR primer.

XX HY2; biliverdin reductase; phytychromobilin synthase; plant;

KW enzyme; PCR; primer; ss.

OS Arabidopsis thaliana.

PN WO200194548-A2.

PD 13-DEC-2001.

PF 05-JUN-2001; 2001WO-US18326.

PR 08-JUN-2000; 2000US-210286P.

PR 26-FEB-2001; 2001US-271758P.

PR 29-MAY-2001; 2001US-0210286.

PA (REGC) UNIV CALIFORNIA.

PI Lagarias JC, Kochi T, Frankenberg N, Gambetta GA, Montgomery BL;

DR WPI; 2002-195566/25.

PT Novel isolated HY2 family bilin reductase having bilin reductase

PT activity, useful for converting biliverdin to phytyobilin, and for

PS producing a photoactive holophytochrome and/or phytyofluor

PS Example 1; Page 49; 102pp; English.

XX The present sequence is that of a primer that was used, with the
 CC primer given in ABA91750, in the PCR amplification of the HY2 locus
 CC on chromosome 3 of Arabidopsis thaliana. The primers amplified a
 CC 1.98 kb fragment of the bacterial artificial chromosome MZB10.18
 CC region of the chromosome from hy2 mutants of Arabidopsis and from
 CC the corresponding wild-type plants. The PCR products were then
 CC sequenced. The HY2 gene (see ABA91766) encodes a ferredoxin-dependent
 CC biliverdin reductase, phytychromobilin synthase (see AAM50863), that
 CC is related to a family of proteins found in oxygenic photosynthetic
 CC bacteria. HY2 is an example of HY bilin reductases of the
 CC invention, which are useful e.g. for the conversion of biliverdin
 CC to phytyobilin and the assembly of holophytochromes or phytyofluors.
 XX

SO Sequence 20 BP; 4 A; 5 C; 4 G; 7 T; 0 other;

Query Match 0.9%; Score 20; DB 24; Length 20;

Best Local Similarity 100.0%; Pred. No. 27;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 CGTTGTCTCAGTAAGTGG 125
 Db 1 CGTTGTCTCAGTAAGTGG 20

RESULT 10

ABA91750/C
 ID ABA91750 standard; DNA; 20 BP.

AC ABA91750;

DT 07-MAY-2002 (first entry)

DE Arabidopsis HY2 locus PCR primer.

XX HY2; biliverdin reductase; phytychromobilin synthase; plant;

KW enzyme; PCR; primer; ss.

OS Arabidopsis thaliana.

PN WO200194548-A2.

PD 13-DEC-2001.

PF 05-JUN-2001; 2001WO-US18326.

PR 08-JUN-2000; 2000US-210286P.

PR 26-FEB-2001; 2001US-271758P.

PR 29-MAY-2001; 2001US-0210286.

PA (REGC) UNIV CALIFORNIA.

PI Lagarias JC, Kochi T, Frankenberg N, Gambetta GA, Montgomery BL;

DR WPI; 2002-195566/25.

PT Novel isolated HY2 family bilin reductase having bilin reductase

PT activity, useful for converting biliverdin to phytyobilin, and for

PS producing a photoactive holophytochrome and/or phytyofluor

PS Example 1; Page 49; 102pp; English.

XX The present sequence is that of a primer that was used, with the
 CC primer given in ABA91749, in the PCR amplification of the HY2 locus
 CC on chromosome 3 of Arabidopsis thaliana. The primers amplified a
 CC 1.98 kb fragment of the bacterial artificial chromosome MZB10.18
 CC region of the chromosome from hy2 mutants of Arabidopsis and from
 CC the corresponding wild-type plants. The PCR products were then
 CC sequenced. The HY2 gene (see ABA91766) encodes a ferredoxin-dependent
 CC biliverdin reductase, phytychromobilin synthase (see AAM50863), that
 CC is related to a family of proteins found in oxygenic photosynthetic
 CC bacteria. HY2 is an example of HY bilin reductases of the
 CC invention, which are useful e.g. for the conversion of biliverdin
 CC to phytyobilin and the assembly of holophytochromes or phytyofluors.
 XX

SO Sequence 20 BP; 8 A; 4 C; 3 G; 5 T; 0 other;

Query Match 0.9%; Score 20; DB 24; Length 20;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2024 TCTGCATTTCAAGATGATG 2043
 Db 20 TCTGCATTTCAAGATGATG 1

RESULT 11

AAC25855
 ID AAC25855 standard; cDNA; 299 BP.

AC AAC5855;
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein 5' EST, SEQ ID NO: 29930.
 XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1033401-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 21-FEB-2000; 2000EP-0200610.
 XX
 PR 26-FEB-1999; 99US-0122487.
 XX
 PA (GEST) GENSET.
 XX
 PI Dumas Milne Edwards J, Duclert A, Giordano J;
 XX
 DR WPI; 2000-500381/45.
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 1; SEQ ID 29930; 71pp + CD-ROM; English.
 XX
 CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.
 CC
 SQ Sequence 299 BP; 100 A; 46 C; 53 G; 99 T; 1 other;
 XX
 XX
 Query Match 0.9%; Score 20; DB 21; Length 299;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 137 TCTCTTTATGATTAAGAA 156
 DB 187 TCTCTTTATGATTAAGAA 206
 XX
 XX
 RESULT 12
 AAS4657
 ID AAS4657 standard; DNA; 8420 BP.
 XX
 AC AAS4657;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Tumour suppressor gene derived chemically modified sequence #379.
 XX
 KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.
 XX
 OS Homo sapiens.
 XX

PN WO200168912-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 15-MAR-2001; 2001WO-EP02955.
 XX
 PR 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2001-602752/68.
 XX
 PT Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 PT cancer
 XX
 PS Claim 1; SEQ ID No 379; 27pp; English.
 XX
 CC The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pcl_sequences.
 CC
 SQ Sequence 8420 BP; 1610 A; 245 C; 2452 G; 4113 T; 0 other;
 XX
 XX
 Query Match 0.9%; Score 20; DB 22; Length 8420;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 492 AGTGTAGTGGAGATTATA 511
 DB 57 AGTGTAGTGGAGATTATA 76
 XX
 XX
 RESULT 13
 ABR31324
 ID ABR31324 standard; DNA; 8420 BP.
 XX
 AC ABR31324;
 XX
 DT 23-APR-2002 (first entry)
 XX
 DE Signal transduction associated gene modified DNA #84.
 XX
 KW Human; signal transduction associated gene; cytosine methylation state;
 KW CpG island; signal transduction associated disease; solid tumour; cancer;
 KW antitumour; cytostatic; mutant; ds.
 KW

XX Homo sapiens.
 OS Synthetic.
 XX WO200200926-A2.
 XX 03-JUN-2002.
 XX 29-JUN-2001; 2001WO-EP07472.
 XX 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIDEMIOLOGICS AG.
 PI Olek A, Piepenbrock C, Berlin K;
 DR WPI; 2002-147896/19.
 XX
 PT Oligonucleotide for diagnosis and therapy of diseases associated with
 PT signal transduction e.g. cancer, comprises chemically modified genomic
 PT sequences of genes associated with signal transduction
 XX
 PS Claim 1; SEQ ID No 167; 24pp; English.
 XX
 CC The present invention relates to chemically modified DNA sequences of
 CC signal transduction associated genes. The DNA sequences are chemically
 CC modified using a solution of bisulphite, hydrogen sulphite or
 CC disulphite. Also disclosed are oligonucleotides and/or PNA oligomers
 CC for detecting the cytosine methylation state (CpG islands) of these
 CC genes, and a method for the diagnosis and/or therapy of genetic and
 CC epigenetic parameters of genes associated with signal transduction.
 CC The genomic DNA can be obtained from cells or cellular components which
 CC contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,
 CC cerebral spinal fluid, tissue embedded in paraffin such as tissue from
 CC eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver,
 CC histologic object slides, and all their possible combinations. The
 CC sequences of the invention are useful for the diagnosis and therapy of
 CC diseases associated with signal transduction e.g. solid tumours and
 CC cancer. ABK31158-ABK31545 represent chemically pretreated genomic DNA
 CC sequences of different genes associated with signal transduction, or
 CC their complementary sequences.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 SO Sequence 8420 BP; 1610 A; 245 C; 2452 G; 4113 T; 0 other;
 Query Match 0.9%; Score 20; DB 24; Length 8420;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 492 AGGTAGTGGAGGATTATTA 511
 Db 57 AGGTAGTGGAGGATTATTA 76
 ||||||||||||||||||
 RESULT 14
 ABN79456/c
 ID ABN79456 standard; cDNA; 399 BP.
 XX
 AC ABN79456;
 XX
 DT 08-JUL-2002 (first entry)
 XX
 DE Human ORF4403 cDNA, SEQ ID NO:8805.
 XX
 KW Human: ORF; open reading frame; ORF; drug screening; diagnosis;
 KW disease monitoring; cytokine; cell proliferation; cell differentiation;
 KW immune modulation; haematopoiesis regulation; tissue growth;
 KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
 KW thrombolytic; tumour inhibition; bodily characteristics; fertility;
 KW behaviour; cancer; proliferative disorder; neurological disorder;

KW cardiovascular disease; immune system disorder; organ transplantation;
 KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
 KW hypothyroidism; cholesterol ester storage disease; infection; vunerary;
 KW vasotropic; antipsoriatic; antidiabetic; cytostatic; neurotic;
 KW neuroprotective; antihypertensive; anticoagulant; thrombolytic;
 KW cardiac; hypotensive; antithyroid; antineoplastic; immunomodulator;
 KW dermatological; analgesic; vitruide; antibacterial; fungicide; gene; ss.
 XX
 OS Homo sapiens.
 XX WO200190366-A2.
 XX 29-NOV-2001.
 XX 24-MAY-2001; 2001WO-US17076.
 XX 24-MAY-2000; 2000US-206630P.
 XX
 PA (CURA-) CURAGEN CORP.
 PI Leach MD, Shinkets RA;
 DR WPI; 2002-106200/14.
 DR P-PSDB; ABP35430.
 XX
 PT Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and disorders related to organ
 PT transplantation
 XX
 PS Claim 1; Page 2444; 2508pp; English.
 XX
 CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins
 CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-
 CC ABN75087 represent cDNAs encoding them. The invention also encompasses
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
 CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies
 CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
 CC polypeptides, methods of screening for modulators of ORFX expression or
 CC activity, and methods of screening individuals for a predisposition to an
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
 CC range of biological activities, such as cytokine, cell proliferation,
 CC cell differentiation, immune modulation, haematopoiesis regulation,
 CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
 CC chemokinetic activity, haemostatic activity, thrombolytic activity,
 CC receptor/ligand, antineoplastic activity, tumour inhibition activity,
 CC and antineoplastic activity, and may also be involved in the determination
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,
 CC nucleic acids and antibodies may be used in the treatment of cancers,
 CC other proliferative disorders such as psoriasis and benign tumours,
 CC neurological disorders such as epilepsy and Alzheimer's disease,
 CC cardiovascular diseases, immune system disorders, disorders related to
 CC organ transplantation, disorders of tissue growth and regeneration,
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
 CC storage disease, and infectious diseases caused by viral, bacterial,
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a
 CC source of primers and probes, in the detection of ORFX genomic sequences
 CC or transcripts, in the identification and cloning of homologous
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
 CC nucleic acids may additionally be used to produce transgenic animals
 CC which may be useful for studying the function and/or activity of ORFX
 CC protein, and in drug screening. The ORFX proteins may also be used as
 CC immunogens to generate specific antibodies, which are used as
 CC diagnosis, treatment and monitoring of ORFX-associated diseases.
 XX
 SO Sequence 399 BP; 115 A; 87 C; 72 G; 124 T; 1 other;
 Query Match 0.9%; Score 19; DB 24; Length 399;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 710 TTTAAATCTCAAGATT 728
 |||||
 Db 267 TTTAAATCTCAAGATT 249

RESULT 15

ABN60565/c
 ID ABN60565 standard; cDNA; 470 BP.

XX
 AC ABN60565;

XX
 DT 28-JUN-2002 (first entry)

XX
 DE Human cancer related polynucleotide SEQ ID NO 532.

XX
 KW Human; cytostatic; gene expression; gene mapping; tissue profiling;
 KW gene therapy; cancer; tumour; gene; ss.

XX
 OS Homo sapiens.

XX
 PN WO200214500-A2.

XX
 PD 21-FEB-2002.

XX
 PF 16-AUG-2001; 2001WO-US25840.

XX
 PR 16-AUG-2000; 2000US-226326P.

XX
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.

XX
 PI Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;
 PI Lamson G, Scott EM, Zhang G, Kassam A, Pot D, Labat I;

XX
 DR WPI; 2002-241905/29.

XX
 PT New nucleic acid for producing a polypeptide, detecting differentially

XX
 PT expressed genes correlated with a cancerous state of a mammalian cell,
 PT and inhibiting tumor growth -

XX
 PS Claim 1; SEQ ID NO 532; 883bp + Sequence Listing; English.

XX
 CC The invention relates to an isolated polynucleotide (ABN27253-ABN33262)

XX
 CC with cytostatic activity. The polynucleotide is used to produce a

XX
 CC polypeptide, to detect differentially expressed genes correlated with a

XX
 CC cancerous state of a mammalian cell and to inhibit tumour growth. The

XX
 CC polynucleotide is used as a probe in mapping and tissue profiling. The

XX
 CC encoded polypeptide and antibodies to the polypeptide can also be used

XX
 CC for therapeutic and diagnostic purposes. The polynucleotide is useful for

XX
 CC gene therapy.

XX
 CC Note: The sequence data for this patent did not form part of the printed

XX
 CC specification, but was obtained in electronic format directly from WIPO

XX
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 470 BP; 149 A; 84 C; 92 G; 145 T; 0 other;

OY 1126 GGTGACCAAGATAC 1144
 |||||

Db 232 GGTGACCAAGATAC 214

Search completed: June 24, 2003, 15:40:51
 Job time : 338 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 15:31:10 ; Search time 80 Seconds
(without alignments)
8280.273 Million cell updates/sec

Title: US-09-870-406a-32

Perfect score: 2160

Sequence: 1 gattcccaacgacacgctg.....tcgtctctactatcaaca 2160

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/pdata/1/lna/5A.COMB.seq:*
2: /cgn2_6/pdata/1/lna/5B.COMB.seq:*
3: /cgn2_6/pdata/1/lna/6A.COMB.seq:*
4: /cgn2_6/pdata/1/lna/6B.COMB.seq:*
5: /cgn2_6/pdata/1/lna/PCrUS.COMB.seq:*
6: /cgn2_6/pdata/1/lna/Dackfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	0.9	731	2	US-08-316-650-15
2	19	0.9	731	5	PCT-US95-02251-15
3	19	0.9	801	4	US-08-558-935-1
4	19	0.9	895	4	US-08-558-935-3
5	19	0.9	2917	2	US-08-624-581-3
6	19	0.9	3681	1	US-08-571-758-3
7	19	0.9	3681	1	US-08-909-984A-3
8	19	0.9	3681	1	US-08-909-984A-3
9	18	0.8	674	3	US-08-578-634C-5
10	18	0.8	674	4	US-09-430-010-5
11	18	0.8	722	3	US-08-480-640A-222
12	18	0.8	722	4	US-08-686-968C-222
13	18	0.8	722	4	US-08-488-237A-222
14	18	0.8	858	4	US-08-998-416-516
15	18	0.8	1133	3	US-08-811-177A-3
16	18	0.8	1307	2	US-08-867-030B-6
17	18	0.8	1307	5	PCT-US95-06119-6
18	18	0.8	1551	1	US-08-161-286-2
19	18	0.8	1662	3	US-08-811-177A-1
20	18	0.8	1797	2	US-08-853-659A-28
21	18	0.8	2669	1	US-08-594-031-91
22	18	0.8	2669	1	US-08-594-031-101
23	18	0.8	3164	4	US-08-686-968C-1
24	18	0.8	3708	1	US-08-185-232A-1
25	18	0.8	3708	1	US-08-416-523-1
26	18	0.8	3708	3	US-08-789-478-1
27	18	0.8	4797	4	US-09-643-597-134

28	18	0.8	8967	2	US-08-853-659A-6	Sequence 6, Appl
29	18	0.8	8967	2	US-08-853-659A-9	Sequence 9, Appl
30	18	0.8	8967	2	US-08-853-659A-64	Sequence 64, Appl
31	18	0.8	8967	2	US-08-853-659A-67	Sequence 67, Appl
32	18	0.8	24701	2	US-08-853-659A-2	Sequence 2, Appl
33	18	0.8	24701	2	US-08-853-659A-3	Sequence 3, Appl
34	18	0.8	24701	2	US-08-853-659A-60	Sequence 60, Appl
35	18	0.8	24701	2	US-08-853-659A-61	Sequence 61, Appl
36	17	0.8	398	1	US-08-417-460-5	Sequence 5, Appl
37	17	0.8	422	1	US-08-417-460-5	Sequence 5, Appl
38	17	0.8	531	4	US-09-404-879A-96	Sequence 96, Appl
39	17	0.8	582	4	US-09-134-001C-1897	Sequence 1897, App
40	17	0.8	607	4	US-08-961-527-331	Sequence 331, App
41	17	0.8	626	4	US-09-328-111-719	Sequence 719, App
42	17	0.8	981	2	US-08-757-653-171	Sequence 171, App
43	17	0.8	981	2	US-08-823-516-74	Sequence 74, Appl
44	17	0.8	981	3	US-08-759-038-110	Sequence 110, App
45	17	0.8	981	3	US-08-758-314-110	Sequence 110, App

ALIGNMENTS

RESULT 1
US-08-316-650-15
; Sequence 15, Application US/08316650
; Patent No. 5942496
; GENERAL INFORMATION:
; APPLICANT: Bonadio, Jeffrey
; APPLICANT: Roessler, Blake J.
; APPLICANT: Goldstein, Steven A.
; APPLICANT: Lin, Mushan
; TITLE OF INVENTION: METHODS AND COMPOSITIONS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/316,650
; FILING DATE: 30-SEP-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/199,780
; FILING DATE: 30-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UMIC:008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 731 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-316-650-15
Query Match 0.9%; Score 19; DB 2; Length 731;
Best local Similarity 100.0%; Pred. No. 9.1;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1216 TTCCCATGGGAGGGAAT 1234
|||||
Db 170 TTCCCATGGGAGGGAAT 188

RESULT 2

PCT-US95-02251-15
Sequence 15, Application PC/TUS9502251
GENERAL INFORMATION:

APPLICANT:
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR STIMULATING BONE
TITLE OF INVENTION: CELLS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: United States of America
ZIP: 77210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
SOFTWARE: Patentin Release #1.0, Version
CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/02251
FILING DATE: CONCURRENTLY HEREWITH
CLASSIFICATION:
Prior Application Data:
APPLICATION NUMBER: US 08/316,650
FILING DATE: 30-SEP-1994

CLASSIFICATION:
APPLICATION NUMBER: US 08/199,780
FILING DATE: 18-FEB-1994

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:

NAME: Parker, David L.
REGISTRATION NUMBER: 32,165

REFERENCE/DOCKET NUMBER: UMIC009P--
TELECOMMUNICATION INFORMATION:

TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679

TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:
LENGTH: 731 base pairs

TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

PCT-US95-02251-15

Query Match
Best Local Similarity 0.9%; Score 19; DB 5; Length 731;
Matches 19; Conservative 100.0%; Pred. No. 9.1;

QY 1216 TTCCCATGGGAGGGAAT 1234
|||||
Db 170 TTCCCATGGGAGGGAAT 188

RESULT 3

US-08-558-935-1
Sequence 1, Application US/08558935
Patent No. 6228637

GENERAL INFORMATION:
APPLICANT: Kasaoka, Keisuke
APPLICANT: Kadotani, Naoto
APPLICANT: Kuwata, Shigeru

APPLICANT: Hayashi, Yumiko
TITLE OF INVENTION: Recombinant Vector, Method for Giving
TITLE OF INVENTION: Immunity Against PVY-T to Potato Plant, and Potato Plant
TITLE OF INVENTION: Having Immunity Against PVY-T
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch and Birch
STREET: 8110 Gatehouse Road, Suite 500 East
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.

ZIP: 22042

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/558,935
FILING DATE:
CLASSIFICATION: 800

Prior Application Data:
APPLICATION NUMBER: US 08/139,157
FILING DATE: 21-OCT-1993

ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.

REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 760-174P

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050

TELEX: 248345

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

LENGTH: 801 base pairs
TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
HYPOHETICAL: NO

ANTI-SENSE: NO
ORIGINAL SOURCE:

ORGANISM: Potato Virus Y-T
US-08-558-935-1

Query Match
Best Local Similarity 0.9%; Score 19; DB 4; Length 801;
Matches 19; Conservative 100.0%; Pred. No. 9.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 255 TATCATGAGATTGGGTT 273
|||||
Db 381 TATCATGAGATTGGGTT 399

RESULT 4

US-08-558-935-3
Sequence 3, Application US/08558935
Patent No. 6228637

GENERAL INFORMATION:
APPLICANT: Kasaoka, Keisuke
APPLICANT: Kadotani, Naoto
APPLICANT: Kuwata, Shigeru
APPLICANT: Hayashi, Yumiko
TITLE OF INVENTION: Recombinant Vector, Method for Giving
TITLE OF INVENTION: Immunity Against PVY-T to Potato Plant, and Potato Plant
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch and Birch
STREET: 8110 Gatehouse Road, Suite 500 East
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.

ZIP: 22042
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/558,935
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/139,157
FILING DATE: 21-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 760-174P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 895 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-558-935-3

Query Match 0.9%; Score 19; DB 4; Length 895;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 255 TATCATGAGTGTGGTT 273
|||||
DB 466 TATCATGAGTGTGGTT 484

RESULT 5
US-08-624-581-3
Sequence 3, Application US/08624581
Patent No. 5807084
GENERAL INFORMATION:
APPLICANT: de Haan, Peter T
TITLE OF INVENTION: Virus resistant or tolerant cells
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: SANDOZ AGRO. INC.
STREET: 975 California Avenue
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: CA94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/624,581
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Marcus-Wyner, Lynn
REGISTRATION NUMBER: 34,869
REFERENCE/DOCKET NUMBER: 137-1088/SG
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-354-3588
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:

LENGTH: 2917 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: ChimERIC sequence
US-08-624-581-3

Query Match 0.9%; Score 19; DB 2; Length 2917;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 255 TATCATGAGTGTGGTT 273
|||||
DB 2370 TATCATGAGTGTGGTT 2388

RESULT 6
US-08-571-758-3/C
Sequence 3, Application US/08571758
Patent No. 5700675
GENERAL INFORMATION:
APPLICANT: Rubin, Gerry M.
APPLICANT: Therrien, Marc
APPLICANT: Chang, Henry C.
APPLICANT: Karim, Felix D.
APPLICANT: Wasserman, David A.
TITLE OF INVENTION: A No. 5700675el Protein Kinase Required for Ras
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/571,758
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: B96-010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3681 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-571-758-3

Query Match 0.9%; Score 19; DB 1; Length 3681;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 779 CAGCAGTCAACATGATG 797
|||||
DB 197 CAGCAGTCAACATGATG 179

RESULT 7
US-08-909-984A-3/C
; Sequence 3, Application US/08909984A
; Patent No. 5747275
; GENERAL INFORMATION:
; APPLICANT: Rubin, Gerry M.
; APPLICANT: Therrien, Marc
; APPLICANT: Chang, Henry C.
; APPLICANT: Karim, Felix D.
; APPLICANT: Massatman, David A.
; TITLE OF INVENTION: A No. 5747275el Protein Kinase Required for Ras
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/909,984A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B96-010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3681 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-909-984A-3
Query Match 0.9%; Score 19; DB 1; Length 3681;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 779 CAGCAGTACAACTGATG 797
Db 197 CAGCAGTACAACTGATG 179
RESULT 8
US-08-909-983-3/C
; Sequence 3, Application US/08909983
; Patent No. 5747288
; GENERAL INFORMATION:
; APPLICANT: Rubin, Gerry M.
; APPLICANT: Therrien, Marc
; APPLICANT: Chang, Henry C.
; APPLICANT: Karim, Felix D.
; APPLICANT: Massatman, David A.
; TITLE OF INVENTION: A No. 5747288el Protein Kinase Required for Ras
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA

ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/909,983
; FILING DATE: 12-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/571,758
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B96-010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3681 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-909-983-3
Query Match 0.9%; Score 19; DB 1; Length 3681;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 779 CAGCAGTACAACTGATG 797
Db 197 CAGCAGTACAACTGATG 179
RESULT 9
US-08-578-634C-5/C
; Sequence 5, Application US/08578634C
; Patent No. 6025163
; GENERAL INFORMATION:
; APPLICANT: Vladimir Shaminin
; APPLICANT: Ethel Michele De Villiers
; APPLICANT: Irene Leisen
; APPLICANT: Harald Zur Hausen
; TITLE OF INVENTION: DNA CODING FOR A PEPTIDE OF A PAPILLOMA
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of The Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/578,634C
; FILING DATE: 26-JAN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B.
; REGISTRATION NUMBER: 30,605

REFERENCE/DOCKET NUMBER: 8484-007
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-4935
TELEFAX: (650)493-5556
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 674 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1.. 672
US-08-578-634C-5

Query Match
Best Local Similarity 100.0%; Score 18; DB 3; Length 674;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1288 TCGTCTACCAAGAAAA 1305
DB 435 TCGTCTACCAAGAAAA 418

RESULT 10
US-09-430-010-5/C
Sequence 5, Application US/09430010
Patent No. 6395512
GENERAL INFORMATION:
APPLICANT: Vladimir Shamanin
APPLICANT: Ethel Michele De Villiers
APPLICANT: Zur Hausen
APPLICANT: Irene Leigh
APPLICANT: Harald Zur Hausen
TITLE OF INVENTION: DNA CODING FOR A PEPTIDE OF A PAPILLOMA
TITLE OF INVENTION: VIRUS MAIN CAPSID PROTEIN AND USE THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of The Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/430.010
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/578,634
FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Abrams, Samuel B.
REGISTRATION NUMBER: 30,605
REFERENCE/DOCKET NUMBER: 8484-007
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-4935
TELEFAX: (650)493-5556
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 674 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA

FEATURE:
NAME/KEY: CDS
LOCATION: 1.. 672
US-09-430-010-5

Query Match
Best Local Similarity 100.0%; Score 18; DB 4; Length 674;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1288 TCGTCTACCAAGAAAA 1305
DB 435 TCGTCTACCAAGAAAA 418

RESULT 11
US-08-480-640A-222
Sequence 222, Application US/08480640A
Patent No. 6033904
GENERAL INFORMATION:
APPLICANT: Cochran, Mark D.
APPLICANT: Junker, David E.
TITLE OF INVENTION: Recombinant Swinepox Virus
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,640A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 222:
SEQUENCE CHARACTERISTICS:
LENGTH: 722 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
US-08-480-640A-222

Query Match
Best Local Similarity 100.0%; Score 18; DB 3; Length 722;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 713 AATCTCAAGATTAGA 730
DB 546 AATCTCAAGATTAGA 563

RESULT 12
US-08-686-968C-222
Sequence 222, Application US/08686968C
Patent No. 6221361
GENERAL INFORMATION:
APPLICANT: Cochran, Mark D.
APPLICANT: Junker, David E.
TITLE OF INVENTION: Recombinant Swinepox Virus

FILE REFERENCE: 39119-H/JML
CURRENT APPLICATION NUMBER: US/08/686,968C
CURRENT FILING DATE: 1996-07-25
NUMBER OF SEQ ID NOS: 231
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 222
LENGTH: 722
TYPE: DNA
ORGANISM: Swinepox virus
US-08-686-968C-222

Query Match 0.8%; Score 18; DB 4; Length 722;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 713 AATCTCAAGATTAGA 730
|||||
DB 546 AATCTCAAGATTAGA 563

RESULT 13
US-08-488-237A-222
Sequence 222, Application US/08488237A
Patent No. 6251403
GENERAL INFORMATION:
APPLICANT: Cochran, Mark D.
APPLICANT: Junker, David E.
TITLE OF INVENTION: Recombinant Swinepox Virus
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,237A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 222:
SEQUENCE CHARACTERISTICS:
LENGTH: 722 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
US-08-488-237A-222

Query Match 0.8%; Score 18; DB 4; Length 722;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 713 AATCTCAAGATTAGA 730
|||||
DB 546 AATCTCAAGATTAGA 563

RESULT 14
US-08-998-416-516/c

Sequence 516, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
APPLICANT: Reibschung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPTII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwalis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 516:
SEQUENCE CHARACTERISTICS:
LENGTH: 858 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG13630P
US-08-998-416-516

Query Match 0.8%; Score 18; DB 4; Length 858;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 647 TTTCAGAAAGTATAGT 664
|||||
DB 716 TTTCAGAAAGTATAGT 699

RESULT 15
US-08-811-177A-3/c
Sequence 3, Application US/0881177A
Patent No. 6025172
GENERAL INFORMATION:
APPLICANT: Danti, Maria
APPLICANT: Catello, Sergio
TITLE OF INVENTION: Isolation and sequencing of
TITLE OF INVENTION: the Hazel FAD2-N Gene
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
STREET: 701-E 535 13th Street, N.W.
CITY: Washington, DC

STATE: USA
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/811,177A
FILING DATE: 04-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sullivan, Michael G
REGISTRATION NUMBER: 35,377
REFERENCE/DOCKET NUMBER: 1729-331
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-783-6040
TELEFAX: 202-783-6031
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1133 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
US-08-811-177A-3

Query Match 0.88; Score 18; DB 3; Length 1133;
Best Local Similarity 100.08; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1443 GTGAGGAGGAGATGAA 1460
|||
DB 104 GTGAGGAGGAGATGAA 87

Search completed: June 24, 2003, 17:19:07
Job time : 82 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 16:43:14 ; Search time 220 Seconds
(without alignments)
14407.467 Million cell updates/sec

Title: US-09-870-406A-32

Perfect score: 2160

Sequence: 1 gaattccccacgtaacgtg.....tcgtctactaatcaaca 2160

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1042519 seqs, 733713590 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PC1_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PC1_NEW_PUB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2160	100.0	2160	US-10-159-901-32	Sequence 32, Appl
2	2160	100.0	2160	US-09-870-406A-32	Sequence 32, Appl
3	28	1.3	36	US-10-159-901-18	Sequence 18, Appl
4	28	1.3	36	US-09-870-406A-18	Sequence 18, Appl
5	22	1.0	22	US-10-159-901-9	Sequence 9, Appl
6	22	1.0	22	US-09-870-406A-9	Sequence 9, Appl
7	22	1.0	30	US-10-159-901-17	Sequence 17, Appl
8	22	1.0	30	US-09-870-406A-17	Sequence 17, Appl
9	20	0.9	20	US-10-159-901-10	Sequence 10, Appl
10	20	0.9	20	US-10-159-901-15	Sequence 15, Appl
11	20	0.9	20	US-10-159-901-16	Sequence 16, Appl
12	20	0.9	20	US-09-870-406A-10	Sequence 10, Appl
13	20	0.9	20	US-09-870-406A-15	Sequence 15, Appl
14	20	0.9	20	US-09-870-406A-16	Sequence 16, Appl
15	19	0.9	526	US-09-796-692-7575	Sequence 7575, Ap
16	19	0.9	526	US-10-040-862-7575	Sequence 7575, Ap
17	19	0.9	549	US-09-938-842A-600	Sequence 600, App
18	19	0.9	1609	US-10-340-583-1	Sequence 1, Appl
19	19	0.9	1609	US-10-340-778-1	Sequence 1, Appl

c 20	19	0.9	1609	9	US-10-340-580-1	Sequence 1, Appl
c 21	19	0.9	1609	9	US-10-340-581-1	Sequence 1, Appl
c 22	19	0.9	1609	9	US-10-340-582-1	Sequence 1, Appl
c 23	19	0.9	1609	9	US-10-340-583-1	Sequence 1, Appl
c 24	19	0.9	14286	10	US-09-070-927A-162	Sequence 162, App
c 25	19	0.9	127197	9	US-09-754-853A-1	Sequence 31, Appl
c 26	18	0.8	28	9	US-10-159-901-31	Sequence 31, Appl
c 27	18	0.8	28	9	US-09-870-406A-31	Sequence 31, Appl
c 28	18	0.8	170	10	US-09-864-761-24010	Sequence 24010, A
c 29	18	0.8	291	10	US-09-294-093B-2204	Sequence 2204, Ap
c 30	18	0.8	297	10	US-09-864-761-27449	Sequence 27449, A
c 31	18	0.8	353	10	US-09-783-590-9844	Sequence 9844, Ap
c 32	18	0.8	401	10	US-09-946-807-198	Sequence 198, App
c 33	18	0.8	401	10	US-09-795-668-198	Sequence 198, App
c 34	18	0.8	401	10	US-09-918-995-6167	Sequence 6167, App
c 35	18	0.8	409	9	US-10-198-846-917	Sequence 917, App
c 36	18	0.8	424	9	US-09-864-761-10807	Sequence 10807, A
c 37	18	0.8	470	10	US-09-918-995-19640	Sequence 19640, A
c 38	18	0.8	486	9	US-09-736-457-1041	Sequence 1041, Ap
c 39	18	0.8	492	9	US-09-902-941-1041	Sequence 1041, Ap
c 40	18	0.8	492	9	US-09-846-626-1041	Sequence 1041, Ap
c 41	18	0.8	492	9	US-10-017-754-1041	Sequence 1041, Ap
c 42	18	0.8	494	9	US-10-076-622-449	Sequence 449, App
c 43	18	0.8	494	9	US-09-551-621-449	Sequence 449, App
c 44	18	0.8	494	9	US-09-604-287A-449	Sequence 449, App
c 45	18	0.8	494	10	US-09-604-287A-449	Sequence 449, App

ALIGNMENTS

RESULT 1
US-10-159-901-32
Sequence 32, Application US/10159901
Publication No. US20030073235A1
GENERAL INFORMATION:
APPLICANT: LAGARIAS, JOHN
APPLICANT: KOICHI, TAKAYUKI
APPLICANT: FRANKENBERG, NICOLE
APPLICANT: GAMBETTA, GREGORY
APPLICANT: MONTGOMERY, BERONDA
TITLE OF INVENTION: LIGHT CONTROLLED GENE EXPRESSION UTILIZING HETEROLOGOUS PHYTOC
FILE REFERENCE: 407T-907731US
CURRENT APPLICATION NUMBER: US/10/159,901
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: 60/294,463
PRIOR FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 57
SOFTWARE: PatentIn version 3.0
SEQ ID NO 32
LENGTH: 2160
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: CDS
LOCATION: (248)..(469)
FEATURE:
NAME/KEY: CDS
LOCATION: (653)..(769)
FEATURE:
NAME/KEY: CDS
LOCATION: (852)..(947)
FEATURE:
NAME/KEY: CDS
LOCATION: (1034)..(1126)
FEATURE:
NAME/KEY: CDS
LOCATION: (1213)..(1344)
FEATURE:
NAME/KEY: CDS
LOCATION: (1419)..(1523)
FEATURE:
NAME/KEY: CDS

LOCATION (1612)..
FEATURE. (1662)
NAME/KEY: CDS
LOCATION: (1743)..
US-10-159-901-32 (1913)

us-09-870-406a-32.01i.rnpb

Query Match
Best Time

Qy	Local Similarity	Score 2160;	DB 9;
Matches 2160; Conservative	100.0%;	100.0%;	
		Pred. No	

	1	GAATTC	CCCACTCAACGTGAGT	0	Mismatches	length 2160
Dd	1			0		

0;
rindels
GAATTCCCCACGTCACGTGGCGATC
61

61 ACCATGACTCGACGCGATGTTGA...GTCGATTCACACGTGGCGGATC...

121 AACTGCTCGACGGATGTGAATTTCATTTGTCGTTGCCAATTCCCTT

[illegible]

181 TGACAAATTTTATCTCTTTATAGATATGATCTTGCCTTTTAACTGCTCTC

181 TGAAGATTTGAGAGAGAGTGTCCAGCTTTTCAGTTTTT

241 TGTGTATG

|||||..GAGGCTTTATCATGAGATTTCG...GGTGTGA
241 TGTGTGAAATG|||.....
|||||..GAGGAGACCTTTGGTTTCAGTTT...

301 AAACCCACCTTATCATGAGCTTGGCTTTCATTTGGTTCATGCTTAA

|||||GTTTCAATGGGTGCATGCTTCAT
301 AAACCGACTCTGCAAGCCCTTA

361 AAAGAAAGATTTCTTCAAGCCCTAATTTTACCGTTGACAT

AAAGAACATTCACGTTAGAG

421 TTTAGAGAAACCTG
TTTACTTAGAGTCCTGCTGTCTG

421 TTTAGAGAAAGCATCGTCTTGACCGTCTG

481 TACATTTTCGTTAGTCTC
|||||TACATCTCCAGGTATATGCAT

481 TACATTTCTGTAAGTAACTAAGTGAGGAGGATTATATATTTCTCAATA
-CCTTCACATCTCCAGGTATATCGAAT

341 GGTAAATTGATTTGCCTTTCCTCAGTGATTTTGG

241 GGTAAATTGATTGAGTCTCTCTGTAATTTC
GTTCTTCTGCTGTGTAATTTC
 607

501 TTATATAGTAAAGCTTCACTT
502 TTTATATAGTAAAGCTTCACTT
503 TTTATATAGTAAAGCTTCACTT

661 TTTATATAAGTAAAGCTTCAGATTTTCTTCTATCAGGAATCTTACTGTTATAGACTGC 50

...GATTTTGTCTTAATCAGCAACT...
...GACGTTCAGGAAGAATA
LACATGACAGACTAGATCC-
|||||
|||||

...AGTACGATGACAGCAGCTAGATGCGTCAAAATGCGTTTTCAGGAAGAATA 660

[illegible]

.....GCGCATAGAGAATGAGACAATGG-720
..GATTAGACTCTTGAGGAGTATC-----
1 GGA-

CCAGTACCAACTGATTGCTTTTAACTTCA 780
GCACCTTTTGGCATTAGAGATGAGACATGCTTTT

.....CCAGGTTTAACTTCA 780
.....CCCATTTCTTACTTCAATTGATG
.....CCCAAACTGATTGCTTAACTCC

-----CCCATTTCTTACTTTCAATGATTGGTA 840
-----CAGTATGTGGTTGTA 840
-----TAGGCTTTGACTTTGGCGGTTA 840
-----TCCTCCGTTTCTTCAATGATTGGTA 840

...GCTTACGTCCTTGACTTGCGCC...
...GGTTTCATGAGCCTGAGTATGA...
...GATTGTGTGTA 840

[illegible]

...ACGTTACAGAGTTGTATTGTAC...
...CTCCATAT 900

...CGTTACATAGTTGTATTTGTAAGT... 960

960 960

100

1

	QY	961	AGTTATCTGGAGTTATCAGCTGCTGTAATGTCACAACTGATGTCATATATTTTACGTAT	1020
Db	961	AGTTATCTGGAGTTATCAGCTGCTGTAATGTCACAACTGATGTCATATATTTTACGTAT	1020	1021
QY	961	AGTTATCTGGAGTTATCAGCTGCTGTAATGTCACAACTGATGTCATATATTTTACGTAT	1020	1021
Db	1021	GTTCCTCTTTAGGACCTTAATCCCTTGCATCAGTTGACCTACAGACGATTTACCAAGA	1080	1081
QY	1021	GTTCCTCTTTAGGACCTTAATCCCTTGCATCAGTTGACCTACAGACGATTTACCAAGA	1080	1081
Db	1081	CAAGTATTTTAAACAAGATATGTCATATATCCAAATATGCTGAGGTGACCAAGAAT	1140	1141
QY	1081	CAAGTATTTTAAACAAGATATGTCATATATCCAAATATGCTGAGGTGACCAAGAAT	1140	1141
Db	1141	ACACCAATTTCTCAATTCGAATGAATCTAATGCTGAGGTGATTAATGATGATCTTGAG	1200	1201
QY	1141	ACACCAATTTCTCAATTCGAATGAATCTAATGCTGAGGTGATTAATGATGATCTTGAG	1200	1201
Db	1201	ATTATATTGACACTTCCCATGAGGAGGAATTTGACGTGATATCCATTAAGTTTTC	1260	1261
QY	1201	ATTATATTGACACTTCCCATGAGGAGGAATTTGACGTGATATCCATTAAGTTTTC	1260	1261
Db	1261	TGCGCTTTGGTGTGATGAGACTAGCTTTCGCTGCTGCAAGAAATTAAGCTTTTTC	1320	1321
QY	1261	TGCGCTTTGGTGTGATGAGACTAGCTTTCGCTGCTGCAAGAAATTAAGCTTTTTC	1320	1321
Db	1321	TCGCGTTTCTAGACTATCATCATGATATATCTACAGGACCAAACTTAAGCTTTTTC	1380	1381
QY	1321	TCGCGTTTCTAGACTATCATCATGATATATCTACAGGACCAAACTTAAGCTTTTTC	1380	1381
Db	1381	GAACCTTTGACTGAGATCTATCATCTTCTCTACAGCAATGCTGATGACAAATC	1440	1441
QY	1381	GAACCTTTGACTGAGATCTATCATCTTCTCTACAGCAATGCTGATGACAAATC	1440	1441
Db	1441	AAGTGGGAGAGATGATGATCACTCTCTCTACAGCAATGCTGATGACAAATC	1500	1501
QY	1441	AAGTGGGAGAGATGATGATCACTCTCTCTACAGCAATGCTGATGACAAATC	1500	1501
Db	1501	ACCTGACATGCGAGACCAAAAGGTGATTCATTTCTCTTGTGTAATTTGATGTTGA	1560	1561
QY	1501	ACCTGACATGCGAGACCAAAAGGTGATTCATTTCTCTTGTGTAATTTGATGTTGA	1560	1561
Db	1561	ACAGACACTGATCTGATTTGTTACAAAGGATATTAATTCCTTTTGTGTAATTTGATGTTGA	1620	1621
QY	1561	ACAGACACTGATCTGATTTGTTACAAAGGATATTAATTCCTTTTGTGTAATTTGATGTTGA	1620	1621
Db	1621	CATGCTCTCTTAAAGATTTAGTGTGAAGCAAGCAAGGATTTAAACATTTATCC	1680	1681
QY	1621	CATGCTCTCTTAAAGATTTAGTGTGAAGCAAGCAAGGATTTAAACATTTATCC	1680	1681
Db	1681	CATTAGTCTCCCATTAATTAATTTAGCTTGTGAAGATGTTGAAATTAATTTATCC	1740	1741
QY	1681	CATTAGTCTCCCATTAATTAATTTAGCTTGTGAAGATGTTGAAATTAATTTATCC	1740	1741
Db	1741	AGAGCTGCTTAAGGATTTCTGTTCAATGGGGTGGATGATGTTGACACAAATCAATC	1800	1801
QY	1741	AGAGCTGCTTAAGGATTTCTGTTCAATGGGGTGGATGATGTTGACACAAATCAATC	1800	1801
Db	1801	TTGATTTCTCTCCAGAGTACCAAAACAGAAATGGAATGTGAACCAATTTATGCAATCA	1860	1861
QY	1801	TTGATTTCTCTCCAGAGTACCAAAACAGAAATGGAATGTGAACCAATTTATGCAATCA	1860	1861
Db	1861	TGGGAATCTCAATTAAGTCTGCTCATTGAGGATTTAAAGCAATTTATGCAATCA	1920	1921
QY	1861	TGGGAATCTCAATTAAGTCTGCTCATTGAGGATTTAAAGCAATTTATGCAATCA	1920	1921
Db	1921	GATATATGTGAAACAGTACATTTCAAGAGTATTAACAGACAAATTTATGCGCTTAACAT	1980	1981
QY	1921	GATATATGTGAAACAGTACATTTCAAGAGTATTAACAGACAAATTTATGCGCTTAACAT	1980	1981
Db	1981	AGTGGATTAAGAAAGACAGCATGAGAGTCTCAAGGCTATCTGCAATTTCAAGATTA	2040	2041
QY	1981	AGTGGATTAAGAAAGACAGCATGAGAGTCTCAAGGCTATCTGCAATTTCAAGATTA	2040	2041
Db	2041	AGTGGATTAAGAAAGACAGCATGAGAGTCTCAAGGCTATCTGCAATTTCAAGATTA	2100	2101
QY	2041	AGTGGATTAAGAAAGACAGCATGAGAGTCTCAAGGCTATCTGCAATTTCAAGATTA	2100	2101

Qy 2041 TTGTTGAGTACCATGATGTAGTTTACAGTACGCTCAGCCCTTATCAAAAT 2100
Db 2041 TTGTTGAGTACCATGATGTAGTTTACAGTACGCTCAGCCCTTATCAAAAT 2100
Qy 2101 GAGATCTCTGAGTATGATGATTTAAGAAAATGATGCTCTACCTAATCAACA 2160
Db 2101 GAGATCTCTGAGTATGATGATTTAAGAAAATGATGCTCTACCTAATCAACA 2160

RESULT 2

US-09-870-406a-32
; Sequence 32, Application US/09870406A
; Publication No. US20030104379A1
; GENERAL INFORMATION:
; APPLICANT: LAGARIAS, JOHN
; APPLICANT: KOICHI, TAKAYUKI
; APPLICANT: FRANKENBERG, NICOLE
; APPLICANT: GAMBETTA, GREGORY
; APPLICANT: MONTGOMERY, BERONDA
; TITLE OF INVENTION: HY2 FAMILY OF BILIN REDUCTASES
; FILE REFERENCE: 407T-907720US
; CURRENT APPLICATION NUMBER: US/09/870,406A
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: 60/271,758
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 60/210,286
; PRIOR FILING DATE: 2000-06-08
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 32
; LENGTH: 2160
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (248)..(469)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (653)..(769)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (852)..(947)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1034)..(1126)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1213)..(1344)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1419)..(1523)
; NAME/KEY: CDS
; LOCATION: (1612)..(1662)
; NAME/KEY: CDS
; LOCATION: (1743)..(1913)
US-09-870-406a-32

Query Match 100.0%; Score 2160; DB 9; Length 2160;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2160; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAATTCCTCCACGTAACGTGACTGCTATTCACAGTGGCGGATGTGGCCCTATAGTTGG 60
Db 1 GAATTCCTCCACGTAACGTGACTGCTATTCACAGTGGCGGATGTGGCCCTATAGTTGG 60
Qy 61 ACATGACTCGAGAGATGTTGAATTCATGCTGTCGCCAATTGCGTTGCTCAGTGA 120
Db 61 ACATGACTCGAGAGATGTTGAATTCATGCTGTCGCCAATTGCGTTGCTCAGTGA 120
Qy 121 AACTGTAAATTTTATCTCTTTTATAGATAAGAAATCTGCTTTTTCAGTTTCACTA 180
Db 121 AACTGTAAATTTTATCTCTTTTATAGATAAGAAATCTGCTTTTTCAGTTTCACTA 180

Db 121 AACTGTAAATTTTATCTCTTTTATAGATAAGAAATCTGCTTTTTCAGTTTCACTA 180
Qy 181 TGAAGAAATTTGAAGAGAGTGTCCGAGAAAGAGACTTTGGTTTCAGTTTGAGAGCT 240
Db 181 TGAAGAAATTTGAAGAGAGTGTCCGAGAAAGAGACTTTGGTTTCAGTTTGAGAGCT 240
Qy 241 TGTGTATGCTTATCAATGAGATTGGGTTTTCATTTGGGTCATGCTTCAAGGACC 300
Db 241 TGTGTATGCTTATCAATGAGATTGGGTTTTCATTTGGGTCATGCTTCAAGGACC 300
Qy 301 AAACCCACCTGCTTAATCTCTGCAAGCCCTAATTAAGATCAATTCACGTTGAGAAGAG 360
Db 301 AAACCCACCTGCTTAATCTCTGCAAGCCCTAATTAAGATCAATTCACGTTGAGAAGAG 360
Qy 361 AAAGAAAGATTTCTACTTAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Db 361 AAAGAAAGATTTCTACTTAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Qy 421 TTTAGAAAGAACCGAAGAAAGATGCTTGAACCTTCACATCTCCAGATATATGCAAT 480
Db 421 TTTAGAAAGAACCGAAGAAAGATGCTTGAACCTTCACATCTCCAGATATATGCAAT 480
Qy 481 TACATTTGTTAGTGTAGTGGAGAGATTAATTCCTCAATGCTTCTGCTGCTGCTGCTGCT 540
Db 481 TACATTTGTTAGTGTAGTGGAGAGATTAATTCCTCAATGCTTCTGCTGCTGCTGCTGCT 540
Qy 541 GGTAAATGATTTGAGTTGCTATTAAGAACCAAAATTAATTAATTAATTAATTAATTAAT 600
Db 541 GGTAAATGATTTGAGTTGCTATTAAGAACCAAAATTAATTAATTAATTAATTAATTAAT 600
Qy 601 TTAATTAAGTAAAGTTGATGATTTGTTTCTAATCAGAAACGTTTCAGAAAGATA 660
Db 601 TTAATTAAGTAAAGTTGATGATTTGTTTCTAATCAGAAACGTTTCAGAAAGATA 660
Qy 661 TAGTAGCATGACAGACTAGATGTAGAACCCGAATTCGAATGCTTGTAAATCTTC 720
Db 661 TAGTAGCATGACAGACTAGATGTAGAACCCGAATTCGAATGCTTGTAAATCTTC 720
Qy 721 AAAGATTAGACTTTGAGGAGATATGCAATATGAGATGAGCAATGCAAGTTTAATCTCA 780
Db 721 AAAGATTAGACTTTGAGGAGATATGCAATATGAGATGAGCAATGCAAGTTTAATCTCA 780
Qy 781 GCAGTACAAACTGATGCTTGTAGTCCCAATTCCTTCAATTAATGATGATGTTGTTGA 840
Db 781 GCAGTACAAACTGATGCTTGTAGTCCCAATTCCTTCAATTAATGATGATGTTGTTGA 840
Qy 841 TCTTCGCTTAGGCTTTGACTTTGCGGGTTTCATGAGCCCTGAGATGATGATCTCCATAT 900
Db 841 TCTTCGCTTAGGCTTTGACTTTGCGGGTTTCATGAGCCCTGAGATGATGATCTCCATAT 900
Qy 901 TCTGTGCTAACCTTTTTCACATCTACCAACGTTAATGTTGATGATGATGATGATGAT 960
Db 901 TCTGTGCTAACCTTTTTCACATCTACCAACGTTAATGTTGATGATGATGATGATGAT 960
Qy 961 AGTTATGCTGAGTATACAGTCTGATTTGTCGAAGCGATGATCAATATTTTCTGAT 1020
Db 961 AGTTATGCTGAGTATACAGTCTGATTTGTCGAAGCGATGATCAATATTTTCTGAT 1020
Qy 1021 GTTCTCTTTAGGAGACCTTAATCTTTGACATCACTGACTGACAGAGATTAACAGA 1080
Db 1021 GTTCTCTTTAGGAGACCTTAATCTTTGACATCACTGACTGACAGAGATTAACAGA 1080
Qy 1081 CAAATATTTATTAACAAGATATGTCATATATACAAATATGCTGAGGTGACACAAAGAT 1140
Db 1081 CAAATATTTATTAACAAGATATGTCATATATACAAATATGCTGAGGTGACACAAAGAT 1140
Qy 1141 ACACCAATTTACTCAATTTGCAAGTAAACCTAATGCTGAGGTAAATGATGATCTTTGAG 1200
Db 1141 ACACCAATTTACTCAATTTGCAAGTAAACCTAATGCTGAGGTAAATGATGATCTTTGAG 1200
Qy 1201 ATTTATTTGACAGACTTTCATGCGGAGGAGAAATGATGCTGATCCATTAAGTTTTC 1260
Db 1201 ATTTATTTGACAGACTTTCATGCGGAGGAGAAATGATGCTGATCCATTAAGTTTTC 1260

```

OY      1261 TCGCCTTGGTGTGAGTACAGTTTTCCTCTACGACAAAGAAAAACATAAGGCTTTGTC 1320
        |||
Db      1261 TCGCCTTGGTGTGAGTACAGTTTTCCTCTACGACAAAGAAAAACATAAGGCTTTGTC 1320
OY      1321 TCTGCGTTCTAGAGTACTATCAGGTATATCTAGAGGCGCAAAAGCTAAGGTTTATG 1380
        |||
Db      1321 TCTGCGTTCTAGAGTACTATCAGGTATATCTAGAGGCGCAAAAGCTAAGGTTTATG 1380
OY      1381 GAACTTGTAGTGAATCATCTCTTCTCTACAGGCGATGGCTTGAGATGACATCC 1440
        |||
Db      1381 GAACTTGTAGTGAATCATCTCTTCTCTACAGGCGATGGCTTGAGATGACATCC 1440
OY      1441 AAGTGAAGGAGAGATGGAACCATCTCATGTGAGAGCAATGTGAGACACACACACT 1500
        |||
Db      1441 AAGTGAAGGAGAGATGGAACCATCTCATGTGAGAGCAATGTGAGACACACACACT 1500
OY      1501 ACCGACATGCGCGACAAAGAGTATTCATTTCTCTTTGTGTATTTGCGATGTTGA 1560
        |||
Db      1501 ACCGACATGCGCGACAAAGAGTATTCATTTCTCTTTGTGTATTTGCGATGTTGA 1560
OY      1561 ACAGACACTATCTCTATTTGTACAAATGATTTGATTTGGTTGAGAGATCTGGA 1620
        |||
Db      1561 ACAGACACTATCTCTATTTGTACAAATGATTTGATTTGGTTGAGAGATCTGGA 1620
OY      1621 CATGCTCTTCTTAAAGATTAGTGTAGTGAACAAAGGCAAGCTATTAAGATTGATCC 1680
        |||
Db      1621 CATGCTCTTCTTAAAGATTAGTGTAGTGAACAAAGGCAAGCTATTAAGATTGATCC 1680
OY      1681 CATTAATGTCCTTATTTAATTTAGCTGTGAAATGTTGAAATGATTTCACAAATC 1740
        |||
Db      1681 CATTAATGTCCTTATTTAATTTAGCTGTGAAATGTTGAAATGATTTCACAAATC 1740
OY      1741 AGGAGCTGTAGAGGATTTCTCTGTCATGGGCTGAGATGAGTGAAGCAAAACATTC 1800
        |||
Db      1741 AGGAGCTGTAGAGGATTTCTCTGTCATGGGCTGAGATGAGTGAAGCAAAACATTC 1800
OY      1801 TTTGATTTCTTCCAGAGTACCAACAGAGATGAGACGTAAGGATTAACGAAATCA 1860
        |||
Db      1801 TTTGATTTCTTCCAGAGTACCAACAGAGATGAGACGTAAGGATTAACGAAATCA 1860
OY      1861 TTGGGAGTGCATATGAAGTCTGTCATGGGATTTTACAGAGCAATTTATGGCTAACAT 1920
        |||
Db      1861 TTGGGAGTGCATATGAAGTCTGTCATGGGATTTTACAGAGCAATTTATGGCTAACAT 1920
OY      1921 GATATATGTGAACAAGTCAAGATTTCAAGATCATCAACAGAGGACGTAAGTGA 1980
        |||
Db      1921 GATATATGTGAACAAGTCAAGATTTCAAGATCATCAACAGAGGACGTAAGTGA 1980
OY      1981 AGTAGGAATTAAGAAAGGACGATGAGAGCTCTCAGGCTATCTGCAATTTCAAGATGA 2040
        |||
Db      1981 AGTAGGAATTAAGAAAGGACGATGAGAGCTCTCAGGCTATCTGCAATTTCAAGATGA 2040
OY      2041 TTGTTGAGTTTACCAGTATGATTTTAAAGAAATGATTCGCTACCTAATCAACA 2100
        |||
Db      2041 TTGTTGAGTTTACCAGTATGATTTTAAAGAAATGATTCGCTACCTAATCAACA 2100
OY      2101 GAGAAATCTCGAGTATGATTTTAAAGAAATGATTCGCTACCTAATCAACA 2160
        |||
Db      2101 GAGAAATCTCGAGTATGATTTTAAAGAAATGATTCGCTACCTAATCAACA 2160

```

RESULT 3
US-10-159-901-18/c

Sequence 18, Application US/10159901

Publication No. US20030073235A1

GENERAL INFORMATION:

APPLICANT: LAGARIAS, JOHN

APPLICANT: KOICHI, TAKAYUKI

APPLICANT: FRANKENBERG, NICOLE

APPLICANT: GABETTA, GREGORY

APPLICANT: MONTGOMERY, BERONDA

TITLE OF INVENTION: LIGHT CONTROLLED GENE EXPRESSION UTILIZING HETEROLOGOUS PHYTOCH

```

FILE REFERENCE: 407T-907731US
CURRENT APPLICATION NUMBER: US/10/159,901
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: 60/294,463
PRIOR FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 57
SOFTWARE: PatentIn version 3.0
SEQ ID NO 18
LENGTH: 36
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: Primer
US-10-159-901-18

Query Match
Best Local Similarity 1.38; Score 28; DB 9; Length 36;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1890 GATTAAACAGACAAATTATCGGCTAAC 1917
        |||
Db      36 GATTAAACAGACAAATTATCGGCTAAC 9

```

RESULT 4
US-09-870-406a-18/c

Sequence 18, Application US/09870406a

Publication No. US20030104379A1

GENERAL INFORMATION:

APPLICANT: LAGARIAS, JOHN

APPLICANT: KOICHI, TAKAYUKI

APPLICANT: FRANKENBERG, NICOLE

APPLICANT: GABETTA, GREGORY

APPLICANT: MONTGOMERY, BERONDA

TITLE OF INVENTION: HY2 FAMILY OF BILIN REDUCTASES

CURRENT APPLICATION NUMBER: US/09/870,406A

CURRENT FILING DATE: 2002-09-04

PRIOR APPLICATION NUMBER: 60/271,758

PRIOR FILING DATE: 2001-02-26

PRIOR APPLICATION NUMBER: 60/210,286

PRIOR FILING DATE: 2000-06-08

NUMBER OF SEQ ID NOS: 57

SOFTWARE: PatentIn version 3.0

SEQ ID NO 18

LENGTH: 36

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Primer

US-09-870-406a-18

Query Match

Best Local Similarity 1.38; Score 28; DB 9; Length 36;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1890 GATTAAACAGACAAATTATCGGCTAAC 1917

Db 36 GATTAAACAGACAAATTATCGGCTAAC 9

US-10-159-901-9

Sequence 9, Application US/10159901

Publication No. US20030073235A1

GENERAL INFORMATION:

APPLICANT: LAGARIAS, JOHN

APPLICANT: KOICHI, TAKAYUKI

APPLICANT: FRANKENBERG, NICOLE

APPLICANT: GABETTA, GREGORY

APPLICANT: MONTGOMERY, BERONDA

TITLE OF INVENTION: LIGHT CONTROLLED GENE EXPRESSION UTILIZING HETEROLOGOUS PHYTOC

FILE REFERENCE: 407T-907731US

```

; CURRENT APPLICATION NUMBER: US/10/159,901
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: 60/294,463
; PRIOR FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 22
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Primer
US-10-159-901-9
```

```

Query Match          1.0%; Score 22; DB 9; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.56;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      762 CAATGACGGTTTAACCTCAGCA 783
Db      1 CAATGACGGTTTAACCTCAGCA 22
```

```

RESULT 6
US-09-870-406a-9
; Sequence 9, Application US/09870406A
; Publication No. US20030104379A1
; GENERAL INFORMATION:
; APPLICANT: LAGARIAS, JOHN
; APPLICANT: KOICHI, TAKAYUKI
; APPLICANT: FRANKENBERG, NICOLE
; APPLICANT: GAMBETTA, GREGORY
; APPLICANT: MONTGOMERY, BERONDA
; TITLE OF INVENTION: HY2 FAMILY OF BILIN REDUCTASES
; FILE REFERENCE: 407T-907720US
; CURRENT APPLICATION NUMBER: US/09/870,406A
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: 60/271,758
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 60/210,286
; PRIOR FILING DATE: 2000-06-08
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 22
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Primer
US-09-870-406a-9
```

```

Query Match          1.0%; Score 22; DB 9; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.56;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      762 CAATGACGGTTTAACCTCAGCA 783
Db      1 CAATGACGGTTTAACCTCAGCA 22
```

```

RESULT 7
US-10-159-901-17
; Sequence 17, Application US/10159901
; Publication No. US20030073235A1
; GENERAL INFORMATION:
; APPLICANT: LAGARIAS, JOHN
; APPLICANT: KOICHI, TAKAYUKI
; APPLICANT: FRANKENBERG, NICOLE
; APPLICANT: GAMBETTA, GREGORY
; APPLICANT: MONTGOMERY, BERONDA
; TITLE OF INVENTION: LIGHT CONTROLLED GENE EXPRESSION UTILIZING HETEROLOGOUS PHYTOCHROME
; FILE REFERENCE: 407T-907731US
; CURRENT APPLICATION NUMBER: US/10/159,901
```

```

; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: 60/294,463
; PRIOR FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Primer
US-10-159-901-17
```

```

Query Match          1.0%; Score 22; DB 9; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.57;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      383 GTCCTGCTGCTGCTGCTATAGG 404
Db      9 GTCCTGCTGCTGCTGCTATAGG 30
```

```

RESULT 8
US-09-870-406a-17
; Sequence 17, Application US/09870406A
; Publication No. US20030104379A1
; GENERAL INFORMATION:
; APPLICANT: LAGARIAS, JOHN
; APPLICANT: KOICHI, TAKAYUKI
; APPLICANT: FRANKENBERG, NICOLE
; APPLICANT: GAMBETTA, GREGORY
; APPLICANT: MONTGOMERY, BERONDA
; TITLE OF INVENTION: HY2 FAMILY OF BILIN REDUCTASES
; FILE REFERENCE: 407T-907720US
; CURRENT APPLICATION NUMBER: US/09/870,406A
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: 60/271,758
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 60/210,286
; PRIOR FILING DATE: 2000-06-08
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Primer
US-09-870-406a-17
```

```

Query Match          1.0%; Score 22; DB 9; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.57;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      383 GTCCTGCTGCTGCTGCTATAGG 404
Db      9 GTCCTGCTGCTGCTGCTATAGG 30
```

```

RESULT 9
US-10-159-901-10/c
; Sequence 10, Application US/10159901
; Publication No. US20030073235A1
; GENERAL INFORMATION:
; APPLICANT: LAGARIAS, JOHN
; APPLICANT: KOICHI, TAKAYUKI
; APPLICANT: FRANKENBERG, NICOLE
; APPLICANT: GAMBETTA, GREGORY
; APPLICANT: MONTGOMERY, BERONDA
; TITLE OF INVENTION: LIGHT CONTROLLED GENE EXPRESSION UTILIZING HETEROLOGOUS PHYTOCHROME
; FILE REFERENCE: 407T-907731US
; CURRENT APPLICATION NUMBER: US/10/159,901
; CURRENT FILING DATE: 2002-05-29
```

```
;; PRIORITY APPLICATION NUMBER: 60/294,463
;; PRIOR FILING DATE: 2001-05-29
;; NUMBER OF SEQ ID NOS: 57
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 10
;; LENGTH: 20
;; TYPE: DNA
;; ORGANISM: Artificial
;; FEATURE:
;; OTHER INFORMATION: Primer
US-10-159-901-10

Query Match          0.9%; Score 20; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1205 ATTTGCAGACTTCCCATGG 1224
Db      20 ATTTGCAGACTTCCCATGG 1

RESULT 10
US-10-159-901-15
;; Sequence 15, Application US/10159901
;; Publication No. US20030073235A1
;; GENERAL INFORMATION:
;; APPLICANT: LAGARIAS, JOHN
;; APPLICANT: KOICHI, TAKAYUKI
;; APPLICANT: FRANKENBERG, NICOLE
;; APPLICANT: GABETTA, GREGORY
;; APPLICANT: MONTGOMERY, BERONDA
;; TITLE OF INVENTION: LIGHT CONTROLLED GENE EXPRESSION UTILIZING HETEROLOGOUS PHYTOCHROME
;; FILE REFERENCE: 407T-907731US
;; CURRENT APPLICATION NUMBER: US/10/159,901
;; CURRENT FILING DATE: 2002-05-29
;; PRIOR APPLICATION NUMBER: 60/294,463
;; PRIOR FILING DATE: 2001-05-29
;; NUMBER OF SEQ ID NOS: 57
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 15
;; LENGTH: 20
;; TYPE: DNA
;; ORGANISM: Artificial
;; FEATURE:
;; OTHER INFORMATION: Primer
US-10-159-901-15

Query Match          0.9%; Score 20; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      106 CGTTGTCTCACTGAACATG 125
Db      1 CGTTGTCTCACTGAACATG 20

RESULT 11
US-10-159-901-16/c
;; Sequence 16, Application US/10159901
;; Publication No. US20030073235A1
;; GENERAL INFORMATION:
;; APPLICANT: LAGARIAS, JOHN
;; APPLICANT: KOICHI, TAKAYUKI
;; APPLICANT: FRANKENBERG, NICOLE
;; APPLICANT: GABETTA, GREGORY
;; APPLICANT: MONTGOMERY, BERONDA
;; TITLE OF INVENTION: LIGHT CONTROLLED GENE EXPRESSION UTILIZING HETEROLOGOUS PHYTOCHROME
;; FILE REFERENCE: 407T-907731US
;; CURRENT APPLICATION NUMBER: US/10/159,901
;; CURRENT FILING DATE: 2002-05-29
;; PRIOR APPLICATION NUMBER: 60/294,463
;; PRIOR FILING DATE: 2001-05-29
;; NUMBER OF SEQ ID NOS: 57
```

```
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 16
;; LENGTH: 20
;; TYPE: DNA
;; ORGANISM: Artificial
;; FEATURE:
;; OTHER INFORMATION: Primer
US-10-159-901-16

Query Match          0.9%; Score 20; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2024 TCTGCATTTCAAGATGATTG 2043
Db      20 TCTGCATTTCAAGATGATTG 1

RESULT 12
US-09-870-406a-10/c
;; Sequence 10, Application US/09870406A
;; Publication No. US20030104379A1
;; GENERAL INFORMATION:
;; APPLICANT: LAGARIAS, JOHN
;; APPLICANT: KOICHI, TAKAYUKI
;; APPLICANT: FRANKENBERG, NICOLE
;; APPLICANT: GABETTA, GREGORY
;; APPLICANT: MONTGOMERY, BERONDA
;; TITLE OF INVENTION: HY2 FAMILY OF BILIN REDUCTASES
;; FILE REFERENCE: 407T-907720US
;; CURRENT APPLICATION NUMBER: US/09/870,406A
;; CURRENT FILING DATE: 2002-09-04
;; PRIOR APPLICATION NUMBER: 60/271,758
;; PRIOR FILING DATE: 2001-02-26
;; PRIOR APPLICATION NUMBER: 60/210,286
;; PRIOR FILING DATE: 2000-06-08
;; NUMBER OF SEQ ID NOS: 57
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 10
;; LENGTH: 20
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Primer
US-09-870-406a-10

Query Match          0.9%; Score 20; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1205 ATTTGCAGACTTCCCATGG 1224
Db      20 ATTTGCAGACTTCCCATGG 1

RESULT 13
US-09-870-406a-15
;; Sequence 15, Application US/09870406A
;; Publication No. US20030104379A1
;; GENERAL INFORMATION:
;; APPLICANT: LAGARIAS, JOHN
;; APPLICANT: KOICHI, TAKAYUKI
;; APPLICANT: FRANKENBERG, NICOLE
;; APPLICANT: GABETTA, GREGORY
;; APPLICANT: MONTGOMERY, BERONDA
;; TITLE OF INVENTION: HY2 FAMILY OF BILIN REDUCTASES
;; FILE REFERENCE: 407T-907720US
;; CURRENT APPLICATION NUMBER: US/09/870,406A
;; CURRENT FILING DATE: 2002-09-04
;; PRIOR APPLICATION NUMBER: 60/271,758
;; PRIOR FILING DATE: 2001-02-26
;; PRIOR APPLICATION NUMBER: 60/210,286
;; PRIOR FILING DATE: 2000-06-08
```

NUMBER OF SEQ ID NOS: 57
SOFTWARE: PatentIn version 3.0
SEQ ID NO 15
LENGTH: 20
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Primer
US-09-870-406A-15

Query Match 0.9%; Score 20; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 106 CGTTGCTCTCACTGAACCTG 125
DB 1 CGTTGCTCTCACTGAACCTG 20

RESULT 14

US-09-870-406A-16/c
Sequence 16, Application US/09870406A
Publication No. US20030104379A1
GENERAL INFORMATION:
APPLICANT: LAGARIAS, JOHN
APPLICANT: KOICHI, TAKAYUKI
APPLICANT: FRANKENBERG, NICOLE
APPLICANT: GAMBETTA, GREGORY
APPLICANT: MONTGOMERY, BERONDA
TITLE OF INVENTION: HY2 FAMILY OF BILIN REDUCTASES
FILE REFERENCE: 407T-907720US
CURRENT APPLICATION NUMBER: US/09/870,406A
CURRENT FILING DATE: 2002-09-04
PRIOR APPLICATION NUMBER: 60/271,758
PRIOR FILING DATE: 2001-02-26
PRIOR APPLICATION NUMBER: 60/210,286
PRIOR FILING DATE: 2000-06-08
NUMBER OF SEQ ID NOS: 57
SOFTWARE: PatentIn version 3.0
SEQ ID NO 16
LENGTH: 20
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Primer
US-09-870-406A-16

Query Match 0.9%; Score 20; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2024 TCTGCATTTCAGATGATG 2043
DB 20 TCTGCATTTCAGATGATG 1

RESULT 15

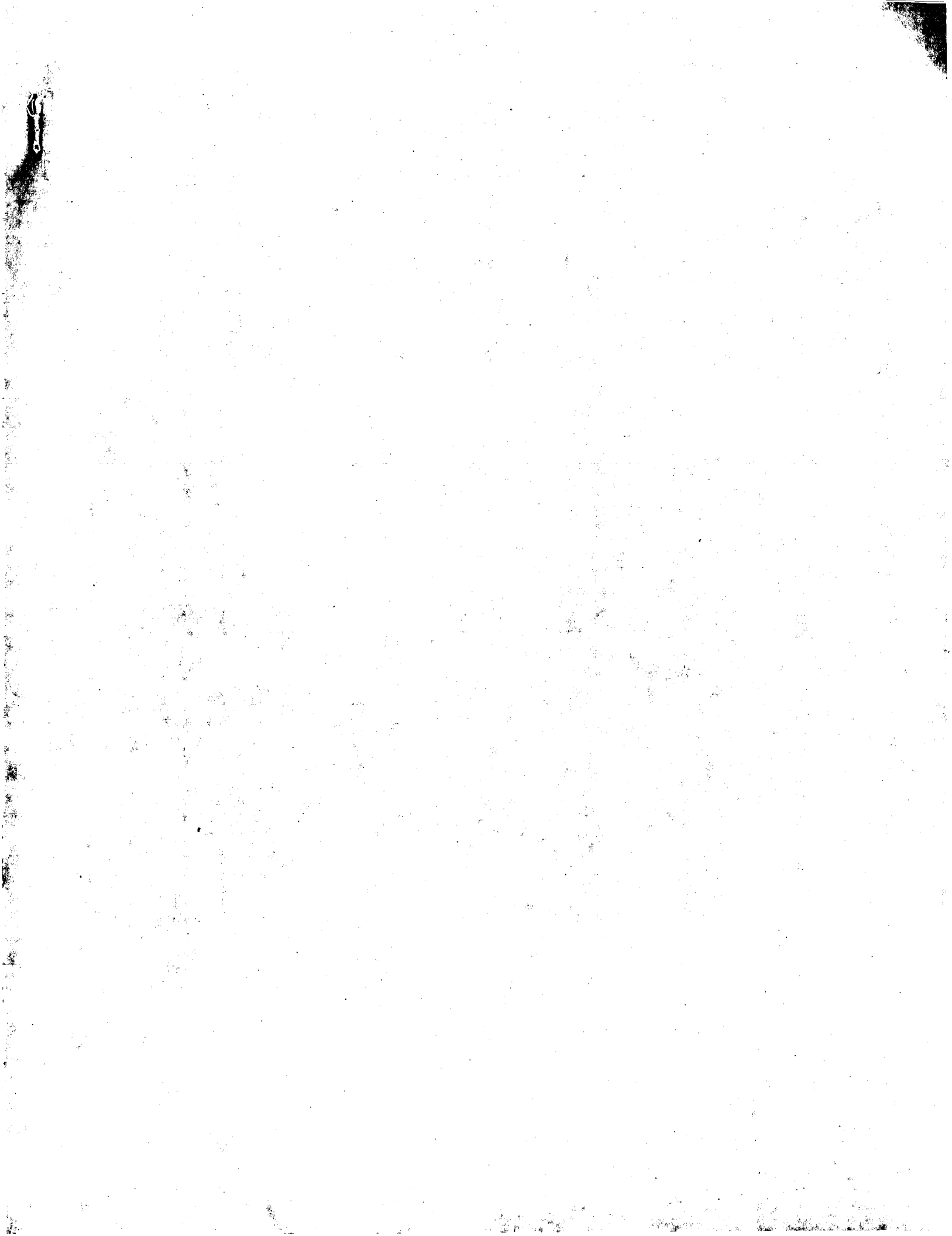
US-09-796-692-7575/c
Sequence 7575, Application US/09796692
Publication No. US20020198362A1
GENERAL INFORMATION:
APPLICANT: Gaiger, Alexander
APPLICANT: Algate, Paul A.
APPLICANT: Mannion, Jane
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
FILE REFERENCE: 2077_001200
CURRENT APPLICATION NUMBER: US/09/796,692
CURRENT FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: 60/186,126
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/190,479
PRIOR FILING DATE: 2000-03-17

PRIOR APPLICATION NUMBER: 60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/202,084
PRIOR FILING DATE: 2000-05-04
PRIOR APPLICATION NUMBER: 60/206,201
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: 60/218,950
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/222,903
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: 60/223,416
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: 60/223,378
PRIOR FILING DATE: 2000-08-07
NUMBER OF SEQ ID NOS: 9597
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 7575
LENGTH: 526
TYPE: DNA
ORGANISM: Homo sapiens
US-09-796-692-7575

Query Match 0.9%; Score 19; DB 9; Length 526;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1300 GAAAAACATTAAGGCTTGT 1318
DB 31 GAAAAACATTAAGGCTTGT 13

Search completed: June 24, 2003, 18:35:29
Job time : 221 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 15:13:05 ; Search time 2053 Seconds
(Without alignments)
17039.577 Million cell updates/sec

Title: US-09-870-406a-32

Perfect score: 2160
Sequence: 1 gaattcccccagcgaacg.....tcgtctctactaatacaca 2160

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size: 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

EST:
1: em_estba:*
2: em_esthm:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	396	18.3	396 10	AV814470 AV814470
2	309	14.3	309 17	CNS00NER
3	264	12.2	417 9	A1999062 701516596
4	111	5.1	111 17	A0011370 F27D23TRC
5	34	1.6	404 12	BG544889 E2953 Ch1
6	26	1.2	285 17	BH543454 BGDJ52TR

c	7	22	1.0	517	17	BH859442	X5-7111b
c	8	22	1.0	684	10	A2420149	A2420149 IM0196N23
c	9	21	1.0	376	10	AM576415	UI-HF-BNO
c	10	21	1.0	392	10	BE525184	M59E65TM
c	11	21	1.0	407	10	AV802211	AV802211
c	12	21	1.0	437	10	AV798190	AV798190
c	13	21	1.0	457	17	A2721635	RPCT-24-7
c	14	21	1.0	472	13	BM489261	pnm2n_pk0
c	15	21	1.0	607	17	BH297378	BH297378
c	16	21	1.0	611	17	A0479208	RPCT-11-2
c	17	21	1.0	720	17	CNS00Y07	AV009075
c	18	20	0.9	177	9	AV009075	AV009075
c	19	20	0.9	243	13	B1023214	CM4-MT024
c	20	20	0.9	356	13	B389702	B389702
c	21	20	0.9	357	10	BE165978	MR3-HT048
c	22	20	0.9	420	9	A1181158	ub95e04.r
c	23	20	0.9	456	13	B370625	B370625
c	24	20	0.9	469	10	BE104555	UI-R-BX0
c	25	20	0.9	483	10	AM175064	FL11h03.Y
c	26	20	0.9	508	17	BH023258	GH-MMB000
c	27	20	0.9	525	17	A2996415	2M0282P06
c	28	20	0.9	528	13	B366249	B366249
c	29	20	0.9	562	12	BF434150	7099005.x
c	30	20	0.9	582	13	B335911	B335911
c	31	20	0.9	591	13	B329755	B329755
c	32	20	0.9	593	17	A0920305	RPCT-23-2
c	33	20	0.9	601	13	B369820	B369820
c	34	20	0.9	619	13	B324742	B324742
c	35	20	0.9	628	13	B332878	B332878
c	36	20	0.9	628	13	B387395	B387395
c	37	20	0.9	632	13	B389891	B389891
c	38	20	0.9	632	13	B1265710	NF083F11T
c	39	20	0.9	660	17	BH077934	RPCT-24-2
c	40	20	0.9	662	14	BQ139097	NF01A11P
c	41	20	0.9	662	17	AG056685	Pan treg1
c	42	20	0.9	665	13	B1265525	NF093C09I
c	43	20	0.9	677	10	AM448757	BRX_1328
c	44	20	0.9	679	12	BG456328	NF077F12P
c	45	20	0.9	686	12	BF639914	NF027C09I

ALIGNMENTS

RESULT 1
AV814470/c
LOCUS
DEFINITION
AV814470 RAF19 Arabidopsis thaliana CDNA clone RAF109-83-B22 3',
mRNA sequence.
ACCESSION
AV814470
VERSION
AV814470.1 GI:19856262
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana
thale cress.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 396)
Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
Arakawa,T., Shibata,K., Shinozaki,A., Muramatsu,M., Hayshtizaki,Y.
and Shinozaki,K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished (2002)
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rkc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI

and XhoI was ligated to modified lambda FLX-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified Bluescript vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further details.

FEATURES

Location/Qualifiers

1..396

/organism="Arabidopsis thaliana"

/db_xref="taxon:3702"

/clone="RAP109-83-B22"

/clone_11b="RAP19"

/dev_stage="plants at various developmental stages from germination to mature seeds"

/lab_host="DH10B"

/note="Site 1: BamHI; Site 2: SalI; subjected to dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24 hr) treatments"

BASE COUNT 111 a 89 c 63 g 133 t

ORIGIN

Query Match 18.3%; Score 396; DB 10; Length 396;

Best Local Similarity 100.0%; Pred. No. 7.1e-202;

Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1752 AGGATTTCTGCTTCATGGGCTGATGATTAGGCAAAAACATTCATTGATTCTT 1811
 DB 396 AGGATTTCTGCTTCATGGGCTGATGATTAGGCAAAAACATTCATTGATTCTT 337
 QY 1812 CCAGATACCAAGAGAGATGAGTAAAGCATTAAGCATATTCATTGAGAGTCA 1871
 DB 336 CCAGATACCAAGAGAGATGAGTAAAGCATTAAGCATATTCATTGAGAGTCA 277
 QY 1872 TATGAACCTGCTCCATGGATTTAAGACAAATTTACGCTAACATGATATATGTA 1931
 DB 276 TATGAACCTGCTCCATGGATTTAAGACAAATTTACGCTAACATGATATATGTA 217
 QY 1932 ACAAGTCAGATTCAGATCATCAACAGAGAGAGTGAAGTGAAGAGATTA 1991
 DB 216 ACAAGTCAGATTCAGATCATCAACAGAGAGAGTGAAGTGAAGAGATTA 157
 QY 1992 GAAAGACCAAGAGAGAGTCTCAAGTCTATCTGATTCGAAGATGTTGATGTT 2051
 DB 156 GAAAGACCAAGAGAGAGTCTCAAGTCTATCTGATTCGAAGATGTTGATGTT 97
 QY 2052 ACCATGATGTGTGTTTAAAGTGAAGTCTCAAGCCTTCATCAAAATGAGATCTCG 2111
 DB 96 ACCATGATGTGTGTTTAAAGTGAAGTCTCAAGCCTTCATCAAAATGAGATCTCG 37
 QY 2112 AGTATGATATGATTTAATGAATGATTTGCTCTC 2147
 DB 36 AGTATGATATGATTTAATGAATGATTTGCTCTC 1

RESULT 2

CONSORDER

LOCUS

DEFINITION 309 bp DNA linear GSS 28-JUN-1999
 Arabidopsis thaliana genome survey sequence SP6 end of BAC F3L24 of IGF library from strain Columbia of Arabidopsis thaliana, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

GENOSCOPE

AL081825
 AL081825.1 GI:5282965
 GSS.
 Arabidopsis thaliana.
 Arabidopsis thaliana.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 309)
 Salanoubat, M., Choise, N., Arliguenave, F., Brottier, P., Winkler, P., Samson, D., Saurin, W., Weissenbach, J. and Quetier, F.
 Unpublished
 2 (bases 1 to 309)
 Genoscope.

TITLE

JOURNAL

Direct Submission
 Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 Evry cedex - FRANCE (E-mail : secref@genoscope.cns.fr)
 Web : www.genoscope.cns.fr

FEATURES

SOURCE

1..309

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db_xref="taxon:3702"

/clone="F3L24"

/clone_11b="IGF"

/note="end : SP6"

BASE COUNT 73 a 56 c 75 g 105 t

ORIGIN

Query Match 14.3%; Score 309; DB 17; Length 309;

Best Local Similarity 100.0%; Pred. No. 5.8e-135;

Matches 309; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CCCAGCTCAACGTGACTGTCATTCACGCTGGCGATGCGCCTATAGTTGACATG 66
 DB 1 CCCAGCTCAACGTGACTGTCATTCACGCTGGCGATGCGCCTATAGTTGACATG 60
 QY 67 ACTCGGAGGATGTTGAATTCATGTCGTTGCCAATTCGTTGTCACAGTGAACGT 126
 DB 61 ACTCGGAGGATGTTGAATTCATGTCGTTGCCAATTCGTTGTCACAGTGAACGT 120
 QY 127 GAAATTTATCTCTTTTATAGATTAAGATCTCTTTTACATTTGCTGATGATGA 186
 DB 121 GAAATTTATCTCTTTTATAGATTAAGATCTCTTTTACATTTGCTGATGATGA 180
 QY 187 AGAATTTGAAGAGAGTCCGAGGAAGAGAGACCTTGTGTTGATGATGATGTTGT 246
 DB 181 AGAATTTGAAGAGAGTCCGAGGAAGAGAGACCTTGTGTTGATGATGATGTTGT 240
 QY 247 AATGCTTATCAATGAGATTTGGTTTCAATGGGTCATGCTTCAAGGACCAACC 306
 DB 241 AATGCTTATCAATGAGATTTGGTTTCAATGGGTCATGCTTCAAGGACCAACC 300
 QY 307 ACCTGTCT 315
 DB 301 ACCTGTCT 309

RESULT 3

LOCUS

DEFINITION 417 bp mRNA linear EST 08-SEP-1999
 thaliana cDNA clone 701516996, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

FEATURES

source

Location/Qualifiers

1. 417
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia Col-0"
 /db_xref="taxon:3702"
 /clone="701516996"
 /clone_lib="A. thaliana, Columbia Col-0, rosette-3"
 /tissue_type="rosette"
 /dev_stage="4 - 7 weeks"
 /note="Vector: pSPORT; Site_1: NotI; Site_2: SalI; cDNA library was derived from untreated rosette tissue from Arabidopsis thaliana, Columbia Col-0, at 4 - 7 weeks. Plants were grown in 1:1:1 peat moss/vermiculite/perlite soil at 22 deg. C +/- 3 deg. C under constant light, and watered with fertilizer. cDNA synthesis was initiated using a NotI-oligo(dT) primer. Double-stranded cDNA was blunt-ended, ligated to SalI adaptors, digested with NotI, size-selected, and cloned into the NotI and SalI sites of the pSPORT vector."

BASE COUNT 112 a 98 c 64 g 143 t
 ORIGIN

Query Match

12.2%; Score 264; DB 9; Length 417;
 Best Local Similarity 99.7%; Pred. No. 1.2e-130;

Matches 384; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1741 AGGAGCTGCTAAGGAGTTCTCTGTCATGCGGTGATGATGAGGACAAACATTC 1800

DB 384 AGGAGCTGCTAAGGAGTTCTCTGTCATGCGGTGATGATGAGGACAAACATTC 335

QY 1801 TTGATTACTTTCAGAGTACCAAGAGATGGAAGTGTAGAGATTAAGAGATTC 1860

DB 324 TTGATTACTTTCAGAGTACCAAGAGATGGAAGTGTAGAGATTAAGAGATTC 265

QY 1861 TTGGGAGTCATATGAAACTGTCATGGATTACAGAGCAATTTATCGGTAAAC 1920

DB 264 TTGGGAGTCATATGAAACTGTCATGGATTACAGAGCAATTTATCGGTAAAC 205

QY 1921 GATATATGTGACAAAGTCAGATTTTCAGATTCACACAAAGAGAGCTAAGTTAGGA 1980

DB 204 GATATATGTGACAAAGTCAGATTTTCAGATTCACACAAAGAGAGCTAAGTTAGGA 146

QY 1981 AGTAGAATTAAGAAAGAGAGCATGAGAGTCTCTGAGTCTTCTGATTTCAAGATGA 2040

DB 145 AGTAGAATTAAGAAAGAGAGCATGAGAGTCTCTGAGTCTTCTGATTTCAAGATGA 86

QY 2041 TTGTTTGTAGTTACCATGATTTAGTTTACAAAGTGTAGCTTCAGCCCTTCACAAAT 2100

DB 85 TTGTTTGTAGTTACCATGATTTAGTTTACAAAGTGTAGCTTCAGCCCTTCACAAAT 26

QY 2101 GAGATCTCGAGTATGATGATTT 2125

DB 25 GAGATCTCGAGTATGATGATTT 1

RESULT 4
 A0011370/c 111 bp DNA linear GSS 29-MAY-1998

LOCUS F27D23TFC IGF Arabidopsis thaliana genomic clone F27D23, DNA

DEFINITION F27D23TFC IGF Arabidopsis thaliana genomic clone F27D23, DNA

ACCESSION A0011370

VERSION A0011370.1 GI:3166615

KEYWORDS GSS.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

REFERENCE Rounsley, S.D., Sub.E.J., Wible, C., Golden, K., Shatsman, S., Choi, P.,
 Yv, K., Akintreloye, B., Shen, K., Gomonsekaram, S., Millscher, J.,
 Adams, M.D. and Venter, J.C.
 A BAC End Sequence Database for Identifying Minimal Overlaps in

JOURNAL

Arabidopsis Genomic Sequencing. Update 4
 Unpublished (1998)
 Other_GSSs: F27D23TFC
 Contact: Steve Rounsley
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: rounsley@tigr.org
 Seq primer: M13 Reverse
 Class: BAC ends
 High quality sequence stop: 111.

FEATURES

Location/Qualifiers

BASE COUNT 29 a 21 c 24 g 37 t
 ORIGIN

Query Match

5.1%; Score 111; DB 17; Length 111;
 Best Local Similarity 100.0%; Pred. No. 3.1e-48;

Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 AAGGACCAACACCACTGTTCTTATCTGCAAGCCCTATATAGATCAATTCACGTTG 352

DB 111 AAGGACCAACACCACTGTTCTTATCTGCAAGCCCTATATAGATCAATTCACGTTG 52

QY 353 AGAAGGAAAGAAAGATTTCTTACTTATAGAGTCTCTGTCGTGCTATAG 403

DB 51 AGAAGGAAAGAAAGATTTCTTACTTATAGAGTCTCTGTCGTGCTATAG 1

RESULT 5

BG544889 404 bp mRNA linear EST 01-MAY-2002

LOCUS

E2953 Chinese cabbage etiolated seedling library Brassica rapa

DEFINITION

subsp. pekinensis cDNA clone E2953, mRNA sequence.

ACCESSION

BG544889

VERSION

BG544889.1 GI:20375870

KEYWORDS

EST.

SOURCE

Brassica rapa subsp. pekinensis

ORGANISM

Brassica rapa subsp. pekinensis

REFERENCE

Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

AUTHORS

Ryu, S.H., Kang, J.S., Kang, C.-h., Kim, C.-Y., Choi, Y.J., Lee, S.-H.,
 Bahk, J.D., Lee, S.-Y., Cho, M.J. and Lim, C.O.
 Expressed Sequence Tags of Chinese Cabbage Etiolated Seedling cDNA
 Unpublished (2001)

TITLE

Contact: Lim, C.O.
 Plant Molecular Biology & Biotechnology Research Centre
 Gyeongsang National University
 #900 Gazwa-dong, Jinju 660-701, Korea
 Tel: 82 55 751 6255
 Fax: 82 55 751 9363
 Email: colim@dongae.gnu.ac.kr
 Seq primer: T7.

FEATURES

Location/Qualifiers

source

1. 404
 /organism="Brassica rapa subsp. pekinensis"
 /cultivar="Jangwon"
 /db_xref="taxon:51351"
 /clone="E2953"
 /clone_lib="Chinese cabbage etiolated seedling library"
 /tissue_type="Etiolated seedling"

```

/lab_host="XL-1 Blue"
/note="Vector: pSPORT 1; site_1: Sal I; site_2: Not I"
BASE COUNT      117 a      94 c      89 g      104 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 34; DB 12; Length 404;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1056 TGACTGACGACGATTACCACAGATTTA 1089
Db      338 TGACTGACGACGATTACCACAGATTTA 371

RESULT 6
BH543454      285 bp      DNA      linear      GSS 14-DEC-2001
LOCUS      BH543454
DEFINITION      BHGDJ52TR BOGD Brassica oleracea genomic clone BHGDJ52, DNA
ACCESSION      BH543454
VERSION      BH543454.1 GI:17795235
KEYWORDS      GSS.
SOURCE      Brassica oleracea.
ORGANISM      Brassica oleracea.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 285)
Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
COMMENT      Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
Location/Qualifiers
1..285
/organism="Brassica oleracea"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone_1fb="BOGDJ52"
/clone_1fb="BOGD"
/note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into PHOS1 using BstXI linkers"
BASE COUNT      71 a      50 c      58 g      106 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 26; DB 17; Length 285;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      867 GGTTCATGAGACCTGAGTATATAC 892
Db      102 GGTTCATGAGACCTGAGTATATAC 127

RESULT 7
BH859442      517 bp      DNA      linear      GSS 08-JUL-2002
LOCUS      BH859442/c
DEFINITION      X5_7111b.t7 Mouse Retroviral Tagged Cancer Gene Database Mus
musculus genomic clone X5_7111b, DNA sequence.
ACCESSION      BH859442
VERSION      BH859442.1 GI:21710263
KEYWORDS      GSS.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

/lab_host="XL-1 Blue"
/note="Vector: pSPORT 1; site_1: Sal I; site_2: Not I"
BASE COUNT      117 a      94 c      89 g      104 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 22; DB 17; Length 517;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      610 TAAAGTTCAGATTTGTTT 631
Db      240 TAAAGTTCAGATTTGTTT 219

RESULT 8
A2420149      684 bp      DNA      linear      GSS 03-OCT-2000
LOCUS      A2420149
DEFINITION      1M0196N23R Mouse 10kb plasmid UOCCIM library Mus musculus genomic
clone UOCCIM0196N23 R, DNA sequence.
ACCESSION      A2420149
VERSION      A2420149.1 GI:10544162
KEYWORDS      GSS.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 684)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
and Wright,D., Weiss,R.
M. Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0196 row: N column: 23
Seq primer: CACACAGAAACAGCTATGAC
Class: plasmid ends
High quality sequence stop: 684.
Location/Qualifiers
1..684
/organism="Mus musculus"
/strain="C57Bl/6J"
FEATURES
SOURCE
REFERENCE
1 (bases 1 to 517)
Suzuki,T., Shen,H., Akagi,K., Morse,H.C., Malley,J.D., Naiman,D.Q.,
Jenkins,N.A. and Copeland,N.G.
Retroviral tagging provides a potent cancer gene discovery tool in
the post-genome-sequence era
Nat Genet., (2002) In press
Contact: Copeland NG
Mouse Cancer Genetics Program
National Cancer Institute
Bldg. 539, Rm. 229, Frederick, MD 21702-1201, USA
Tel: 301 846 1260
Fax: 301 846 6666
Email: copeland@ncicrf.gov
Class: PCR with specific primers.
Location/Qualifiers
1..517
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_1fb="X5_7111b"
/clone_1fb="Mouse Retroviral Tagged Cancer Gene Database"
/sex="female"
/tissue_type="Leukemia"
/note="Inverse PCR method
(http://genome2.ncicrf.gov/RTGCD)"
BASE COUNT      130 a      142 c      106 g      139 t
ORIGIN

REFERENCE
1 (bases 1 to 517)
Suzuki,T., Shen,H., Akagi,K., Morse,H.C., Malley,J.D., Naiman,D.Q.,
Jenkins,N.A. and Copeland,N.G.
Retroviral tagging provides a potent cancer gene discovery tool in
the post-genome-sequence era
Nat Genet., (2002) In press
Contact: Copeland NG
Mouse Cancer Genetics Program
National Cancer Institute
Bldg. 539, Rm. 229, Frederick, MD 21702-1201, USA
Tel: 301 846 1260
Fax: 301 846 6666
Email: copeland@ncicrf.gov
Class: PCR with specific primers.
Location/Qualifiers
1..517
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_1fb="X5_7111b"
/clone_1fb="Mouse Retroviral Tagged Cancer Gene Database"
/sex="female"
/tissue_type="Leukemia"
/note="Inverse PCR method
(http://genome2.ncicrf.gov/RTGCD)"
BASE COUNT      130 a      142 c      106 g      139 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 22; DB 17; Length 517;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      610 TAAAGTTCAGATTTGTTT 631
Db      240 TAAAGTTCAGATTTGTTT 219

RESULT 8
A2420149      684 bp      DNA      linear      GSS 03-OCT-2000
LOCUS      A2420149
DEFINITION      1M0196N23R Mouse 10kb plasmid UOCCIM library Mus musculus genomic
clone UOCCIM0196N23 R, DNA sequence.
ACCESSION      A2420149
VERSION      A2420149.1 GI:10544162
KEYWORDS      GSS.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 684)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
and Wright,D., Weiss,R.
M. Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0196 row: N column: 23
Seq primer: CACACAGAAACAGCTATGAC
Class: plasmid ends
High quality sequence stop: 684.
Location/Qualifiers
1..684
/organism="Mus musculus"
/strain="C57Bl/6J"
FEATURES
SOURCE
REFERENCE
1 (bases 1 to 517)
Suzuki,T., Shen,H., Akagi,K., Morse,H.C., Malley,J.D., Naiman,D.Q.,
Jenkins,N.A. and Copeland,N.G.
Retroviral tagging provides a potent cancer gene discovery tool in
the post-genome-sequence era
Nat Genet., (2002) In press
Contact: Copeland NG
Mouse Cancer Genetics Program
National Cancer Institute
Bldg. 539, Rm. 229, Frederick, MD 21702-1201, USA
Tel: 301 846 1260
Fax: 301 846 6666
Email: copeland@ncicrf.gov
Class: PCR with specific primers.
Location/Qualifiers
1..517
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_1fb="X5_7111b"
/clone_1fb="Mouse Retroviral Tagged Cancer Gene Database"
/sex="female"
/tissue_type="Leukemia"
/note="Inverse PCR method
(http://genome2.ncicrf.gov/RTGCD)"
BASE COUNT      130 a      142 c      106 g      139 t
ORIGIN

```

/db_xref="taxon:10090"
/clone="UUCG1K0196N23"
/clone_lib="Mouse 10kb plasmid UUCG1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g11473211419b1Af129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 216 a 168 c 142 g 158 t
ORIGIN

Query Match 1.0%; Score 22; DB 17; Length 684;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 351 TGAGAGGAGAGAGAGAGATT 372
|||||
DB 271 TGAGAGGAGAGAGAGAGATT 292

RESULT 9
AM576415/c 376 bp mRNA linear EST 15-MAR-2000
LOCUS UT-HF-BNO-alp-a-10-0-UT.s1 NIH_MGC_50 Homo sapiens cDNA clone
DEFINITION IMAGE:3078834 3, mRNA sequence.

ACCESSION AM576415
VERSION AM576415.1 GI:7247954
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 376)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.

Email: cgapbs-rt@mail.nih.gov

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

CDNA Library Preparation: M.B. Soares Lab

CDNA Library Arrayed by: M.B. Soares Lab

DNA Sequencing by: M.B. Soares Lab

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/MLN at:

www-bio.lhnl.gov/bbrp/image/image.html

Seq primer: M13 Forward

POLYA-yes.

FEATURES

source location/Qualifiers

1..376
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone IMAGE:3078834
/clone_id="NIH_MGC_50"
/tissue_type="lymph"

/cell_type="germinal center B cells"
/cell_line="MGC85"
/lab_host="DH10B (LIT)"
/note="Vector: pT73-Pac; Site_1: NotI; Site_2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(3.5-4.4kb). Directionally cloned. Cells provided by
Louis M. Staudt, Ph.D. Library preparation by Maria de
Fatima Bonaldo, Ph.D. and M. Benito Soares, Ph.D."

BASE COUNT 135 a 67 c 54 g 120 t
ORIGIN

Query Match 1.0%; Score 21; DB 10; Length 376;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 259 AATGAGTTGGGTTTCAT 279
|||||
DB 296 AATGAGTTGGGTTTCAT 276

RESULT 10
BE525184/c 392 bp mRNA linear EST 19-MAR-2001
LOCUS M59E6STM Arabidopsis developing seed Arabidopsis thaliana cDNA
DEFINITION clone M59E6 5', mRNA sequence.

ACCESSION BE525184
VERSION BE525184.1 GI:9783162
KEYWORDS EST.
SOURCE thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 392)

White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de

Ilarduya, O., Javoroski, J.G., Ohlrogge, J. and Banning, C.

A new set of Arabidopsis expressed sequence tags from developing

seeds. The metabolic pathway from carbohydrates to seed oil

Plant Physiol. 124 (4), 1582-1594 (2000)

JOURNAL

COMMENT

Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
USA
Tel: 517 355 1609
Fax: 517 353 9334

Email: benning@msu.edu

Michigan State University DNA Sequencing Facility Arabidopsis

Biological Resource Center, The Ohio State University, 309 Botany &

Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:

6142920603 TEL: 6142929371.

Location/Qualifiers

1..392

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db_xref="taxon:3702"

/clone="M59E6"

/clone_lib="Arabidopsis developing seed"

/tissue_type="seed"

/dev_stage="5-13 days after flowering"

/lab_host="E.coli"

/note="Organ: Developing seed; Vector: pBluescript SK-";

Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 105 a 118 c 34 g 135 t
ORIGIN

Query Match 1.0%; Score 21; DB 10; Length 392;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1445 GAGGAGAGATGAGACCAT 1465
|||||

Db 234 GAGGAGAGATGGAACCATC 214

RESULT 11
AV802211/c
LOCUS
DEFINITION
AV802211 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-31-B20 3',
mRNA sequence.

ACCESSION
AV802211
VERSION
AV802211.1 GI:19836196
KEYWORDS
EST.
SOURCE
thale cress.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi
1 (bases 1 to 407)
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,
Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.
and Shinozaki, K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished (2002)
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rcc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified Lambda FIC-1 vector (Carninci et
al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified pluscript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/planl/index_e.html) for further
details.

FEATURES
source
1. Location/Qualifiers
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="RAFL09-31-B20"
/dev_stage="RAFL9"
/dev_stage="plants at various developmental stages from
germination to mature seeds"
/lab_host="DH10B"
/note="Site_1: BamHI; Site_2: SalI; subjected to
dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
hr) treatments"

BASE COUNT
110 a 124 c 35 g 138 t

ORIGIN

Query Match 1.0%; Score 21; DB 10; Length 407;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1445 GAGGAGAGATGGAACCATC 1465
Db 234 GAGGAGAGATGGAACCATC 214

RESULT 12
AV798190/c
LOCUS
DEFINITION
AV798190 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-15-F01 3',
mRNA sequence.

ACCESSION
AV798190
VERSION
AV798190.1 GI:19832173
KEYWORDS
EST.
SOURCE
thale cress.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

REFERENCE
AUTHORS
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 437)
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,
Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.
and Shinozaki, K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished (2002)
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rcc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified Lambda FIC-1 vector (Carninci et
al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified pluscript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/planl/index_e.html) for further
details.

FEATURES
source
1. Location/Qualifiers
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="RAFL09-15-F01"
/dev_stage="RAFL9"
/dev_stage="plants at various developmental stages from
germination to mature seeds"
/lab_host="DH10B"
/note="Site_1: BamHI; Site_2: SalI; subjected to
dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
hr) treatments"

BASE COUNT
109 a 132 c 43 g 153 t

ORIGIN

Query Match 1.0%; Score 21; DB 10; Length 437;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1445 GAGGAGAGATGGAACCATC 1465
Db 234 GAGGAGAGATGGAACCATC 214

RESULT 13
A2721635/c
LOCUS
DEFINITION
A2721635
DNA sequence.

ACCESSION
A2721635
VERSION
A2721635.1 GI:12464522
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 457)
Zhao, S., Nieman, W., Malek, J., Shatsman, S., Akhmet, B., Levin, M.,
Tsegaye, G., Geier, K., Krol, M., Shvartsbeyn, A., Gebregiorgis, E.,
Russell, D., de Jong, P. and Fraser, C.M.
Mouse BAC End Sequences from library RPCI-24
Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html. Plate: 73 row: A column: 18
Seq primer: T7
Class: BAC ends.

FEATURES
source
Location/Qualifiers
1. 457

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-73A18"
/clone_id="RPCI-24"
/sex="Male"

/cell_type="Spleen/Brain"
/note="Vector: pTRABAC1; Site_1: BamHI; Site_2: BamHI; RPCT-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTRABAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT
ORIGIN
172 a 81 c 77 g 127 t

Query Match
Best Local Similarity 100.0%; Pred. No. 16;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY
613 AAGTTCAGATTTGTTTCT 633
|||||
274 AAGTTCAGATTTGTTTCT 254

RESULT 14
LOCUS
BM489261 472 bp mRNA linear EST 07-FEB-2002
DEFINITION
pgm2n.pk010.g14 Normalized Chicken Breast Muscle, Leg Muscle, and Epiphyseal Growth Plate cDNA library (pgm2n) Gallus gallus cDNA clone pgm2n.pk010.g14 5' similar to no significant hits (plog(p)

4), mRNA sequence.

ACCESSION
BM489261
VERSION
BM489261.1 GI:18610192
KEYWORDS
EST.
SOURCE
Chicken.
ORGANISM
Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianidae; Gallus.
1 (bases 1 to 472)
Coburn, L.A. and Monsonego-Ornan, E.
ESTs from Normalized Chicken Breast Muscle, Leg Muscle, and Epiphyseal Growth Plate cDNA library, USDA/TRAIS Animal Genome Project

REFERENCE
AUTHORS
TITLE
Epiphyseal Growth Plate cDNA library, USDA/TRAIS Animal Genome Project
JOURNAL
Unpublished (2002)
COMMENT
Contact: Larry A. Coburn
University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822
Email: coburn@udel.edu, www.chickest.udel.edu.
Location/Qualifiers
1. 472

FEATURES
source
Location/Qualifiers
1. 472

/organism="Gallus gallus"
/strain="Commercial broiler and Ottawa Res. Centre
Strains 90 & 21"
/db_xref="taxon:9031"
/clone="pgm2n.pk010.g14"
/clone_id="Normalized Chicken Breast Muscle, Leg Muscle, and Epiphyseal Growth Plate cDNA library (pgm2n)"
/sex="Male and Female"
/tissue_type="Breast muscle, leg muscle and epiphyseal growth plate"
/dev_stage="Breast, leg, embryo(d19); post-hatch(d1,1,3,5,7,9,11 weeks); growth plate(d1d,7d,14d post-hatch)"

/lab_host="E. coli EMDH10B"
/note="Vector: pCMVSPORT6; Library made from equivalent pools of total RNA isolated from each tissue (embryonic muscle 33.3%, juvenile muscle 33.3%, and epiphyseal growth plate 33.3% of the final RNA pool). Single pass sequencing from 5'-end"

BASE COUNT
ORIGIN
140 a 65 c 87 g 171 t 9 others

Query Match
Best Local Similarity 100.0%; Pred. No. 16;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY
2116 TGATATGATTTTATGAAAT 2136
|||||
288 TGATATGATTTTATGAAAT 308

RESULT 15
LOCUS
BH297378/c 607 bp DNA linear GSS 30-NOV-2001
DEFINITION
CH230-45E12.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
CH230-45E12, DNA sequence.

ACCESSION
BH297378
VERSION
BH297378.1 GI:17209786
KEYWORDS
GSS.
SOURCE
Norway rat.
Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 607)

REFERENCE
AUTHORS
Zhao, S., Shetty, J., Shatsman, S., Tesgaye, C., Geer, K., Shvartsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de Jong, P. and Fraser, C.M.
Rat BAC End sequences from Library CHORI-230 EcORI segment
Unpublished (1999)
Other GSSs: CH230-45E12-TV
JOURNAL
COMMENT
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 0208

Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230 (<http://www.choi.org/bacpac/rat230.htm>). For BAC library availability, please contact Pieter de Jong (pdejong@mail.choi.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html. Plate: 45 row: E column: 12
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
1. 607

FEATURES
source
Location/Qualifiers
1. 607
/organism="Rattus norvegicus"
/strain="BN/SSNhsd/MKW"
/db_xref="taxon:10116"
/clone="CH230-45E12"
/clone_id="CHORI-230 Segment 1"
/sex="Female"
/cell_type="Brain"
/note="Vector: pTRABAC2.1; Site_1: EcoRI; Site_2: EcoRI; CHORI-230 Rat (BN/SSNhsd/MKW) BAC library produced by Pieter de Jong"

BASE COUNT
ORIGIN
139 a 133 c 153 g 182 t

Query Match
Best Local Similarity 100.0%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1106 ATATATCACAATATGCTGAG 1126
|||||
Db 281 ATATATCACAATATGCTGAG 261

Search completed: June 24, 2003, 17:17:32
Job time : 2057 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 18:53:41 ; Search time 17 Seconds

(Without alignments)
1860.484 Million cell updates/sec

Title: US-09-870-406A-33

Perfect score: 329
Sequence: 1 MALSMERGFSGSCFKAPNP.....SIIGKSYETRPMDLTGQFIG 329Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	2.7	476	2 A46118	myosin-binding pro
2	8	2.4	371	2 F90486	dehydrogenase, pro
3	8	2.4	491	2 C86390	hypothetical prote
4	8	2.4	786	2 C86406	88.6k hypothetical
5	8	2.4	944	2 S66870	DNA4 protein - yea
6	7	2.1	120	2 T44411	ribosomal protein
7	7	2.1	126	2 G70616	hypothetical prote
8	7	2.1	158	2 D71111	hypothetical prote
9	7	2.1	160	2 S43632	H+-transporting tw
10	7	2.1	160	1 VCTNS	coat protein - tob
11	7	2.1	207	2 C70909	probable two-compo
12	7	2.1	229	2 T18109	ankyrin repeat pro
13	7	2.1	256	2 A10267	probable membrane
14	7	2.1	262	2 T34691	hypothetical prote
15	7	2.1	283	2 C70478	protease IV - Ag
16	7	2.1	287	2 AC1805	B. subtilis SpoIII
17	7	2.1	287	2 AE1431	B. subtilis SpoIII
18	7	2.1	289	2 D69825	glucose 1-dehydrog
19	7	2.1	294	2 C82497	probable ABC transp
20	7	2.1	304	2 H75610	probable agmatinas
21	7	2.1	309	2 T01255	probable ethylene
22	7	2.1	313	2 G95883	probable ABC trans
23	7	2.1	315	2 T06053	phosphate ubiquitin
24	7	2.1	324	2 A70584	gene 33 protein -
25	7	2.1	330	1 Q0B8X2	gppI protein - The
26	7	2.1	344	2 D72321	apolipoprotein H P
27	7	2.1	345	1 NBMS	hypothetical prote
28	7	2.1	346	2 AD2242	steroid/chyroid/re
29	7	2.1	356	2 T34133	

30	7	2.1	362	2 T19031	hypothetical prote
31	7	2.1	363	2 D81037	GTP-binding protei
32	7	2.1	371	2 D84186	hypothetical prote
33	7	2.1	385	2 H86203	hypothetical prote
34	7	2.1	387	2 F83577	phosphoglycerate k
35	7	2.1	389	2 T20604	hypothetical prote
36	7	2.1	392	2 B69321	cell division prot
37	7	2.1	408	2 S76830	hypothetical prote
38	7	2.1	433	2 F69458	coenzyme F390 synt
39	7	2.1	439	2 A97288	uncharacterized co
40	7	2.1	456	2 D95384	protein [imported
41	7	2.1	465	2 I56523	GABA-A receptor rh
42	7	2.1	465	2 A38079	gamma-aminobutyric
43	7	2.1	467	2 T15892	hypothetical prote
44	7	2.1	468	2 H72230	glutamate synthase
45	7	2.1	472	2 D71076	probable glutamate

ALIGNMENTS

RESULT 1

A46118 myosin-binding protein H - human

C:Species: Homo sapiens (man)

C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 28-May-1999

C:Accession: A46118

R:Vaughan, K.T.; Weber, F.E.; Ried, T.; Ward, D.C.; Relnach, F.C.; Fischman, D.A.

Genomics 16, 34-40, 1993

A:Title: Human myosin-binding protein H (MyBP-H): complete primary sequence, genomic

A:Reference number: A46118; MUID:93252409; PMID:8486381

A:Accession: A46118

A>Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 1-476 <VAV>

A:Cross-references: GB:L05606; NID:9292336; PIDN:AAA6339.1; PID:9292337

A:Experimental source: striated muscle

A>Note: sequence extracted from NCBI backbone (NCBI:131208, NCBI:P131209)

Query Match	2.7%	Score 9;	DB 2;	Length 476;
Best Local Similarity	100.0%;	Pred. No. 0.5;		
Matches	9;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;

RESULT 2

F90486 dehydrogenase, probable [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001

C:Accession: F90486

R:She, O.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayer, M.J.; Ch

Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thl-Hgoc, H.P.; Redder

arrett, R.A.; Ragan, M.A.; Sersen, C.W.; Van der Oost, J.

Submitted to Genbank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: F90486

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-371 <KUR>

A:Cross-references: GB:AE006641; NID:913816451; PIDN:AAK43149.1; GSPDB:GN00155

C:Genetics:

A:Gene: SSO3049

Query Match	2.4%	Score 8;	DB 2;	Length 371;
Best Local Similarity	100.0%;	Pred. No. 4.6;		
Matches	8;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;

OY 272 KELLRDL 279

DB 318 KELLDFLN 325

RESULT 3

Hypothetical protein TIK7_26 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C:Accession: C66390
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Jensen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A66141; MUID:21016719; PMID:11130712
A:Accession: C66390
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-491 <STO>
A:Cross-references: GB:AE005172; NID:99797761; PIDN:AAF98579.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match

Best Local Similarity 100.0%; Score 8; DB 2; Length 491;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 274 LNRDFLN 281
DB 290 LNRDFLN 297

RESULT 4

C66406
88.6K hypothetical protein F13K9.16 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Nov-2001
C:Accession: C66406
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Jensen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A66141; MUID:21016719; PMID:11130712
A:Accession: C66406
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-786 <STO>
A:Cross-references: GB:AE005172; NID:911024878; PIDN:AMC26962.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match

Best Local Similarity 100.0%; Score 8; DB 2; Length 786;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 GEAKAKEL 274
DB 448 GEAKAKEL 455

RESULT 5

S66870
DNL4 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein 02525; protein UND407; protein UNE452; protein YOR005C
C:Species: Saccharomyces cerevisiae
C>Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 29-Oct-1999
C:Accession: S66870; S61988; S61989; S72137; S72138
R:Patterson, B.; Sterky, F.; Uhlen, M.
Submitted to the Protein Sequence Database, July 1996
A:Reference number: S66882
A:Accession: S66870
A:Molecule type: DNA
A:Residues: 1-944 <P>

A:Cross-references: EMBL:274913; NID:91420095; PIDN:CAA99193.1; PID:e252317; PID:9142
A:Experimental source: strain S288C
R:Sterky, F.; Uhlen, M.
Submitted to the EMBL Data Library, December 1995
A:Reference number: S61981
A:Accession: S61988
A:Molecule type: DNA
A:Residues: 493-944 <STO>

A:Cross-references: EMBL:U43491; NID:91150992; PIDN:AA49484.1; PID:91151000
A:Accession: S61989
A:Molecule type: DNA
A:Residues: 1-407 <STA>
A:Cross-references: EMBL:U43491; NID:91150992; PIDN:AA49485.1; PID:91151001
R:Sterky, F.; Holmberg, A.; Fellersson, B.; Uhlen, M.
Yeast 12, 1091-1095, 1996
A:Title: The sequence of a 30 kb fragment on the left arm of chromosome XV from Sacch

A:Reference number: S72137
A:Accession: S72138
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 493-944 <STW>

A:Cross-references: EMBL:U43491; NID:91150992; PIDN:AA49484.1; PID:91151000
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1995
A:Accession: S72138
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-407 <STP>

A:Cross-references: EMBL:U43491; NID:91150992; PIDN:AA49485.1; PID:91151001
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1995
C:Genetics:
A:Gene: SGD:DNA4
A:Cross-references: SGD:S0005531; MIPS:YOR005C
A:Map position: 15R

Query Match
Best Local Similarity 100.0%; Score 8; DB 2; Length 944;
Matches 8; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 186 GESIKFFS 193
DB 296 GESIKFFS 303

RESULT 6

T44411
ribosomal protein L17 rplO [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 15-Jun-2001
C:Accession: T44411; C83670
R:Takami, H.; Takaki, Y.; Nakasone, K.; Hirama, C.; Inoue, A.; Horikoshi, K.
Biosci. Biotechnol. Biochem. 63, 452-455, 1999
A:Title: Sequence analysis of a 32-kb region including the major ribosomal protein ge
A:Reference number: 222756; MUID:99209008; PMID:10192928
A:Accession: T44411
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-120 <TAK>
A:Cross-references: EMBL:AB017508; NID:94512395; PIDN:BAA75299.1; PID:94512432
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
 A:Reference number: A83650; MUID:20512582; PMID:11058132
 A:Accession: C83670
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-120 <STO>
 A:Cross-references: GB:AP001507; GB:BA000004; NID:910172612; PIDN:BA03882.1; GSPDB:GNOC
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: rplQ
 C:Superfamily: Escherichia coli ribosomal protein L17

Query Match 2.1%; Score 7; DB 2; Length 120;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 268 EAKAKEL 274
 |||||
 DB 34 EAKAKEL 40

RESULT 7

hypothetical protein RV0140 - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: G70616
 R:Cole, S.T.; Brosch, R.; Parhill, J.; Gartner, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, K.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98255987; PMID:9654220
 A:Accession: G70616
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-126 <COL>
 A:Cross-references: GB:292770; GB:AL123456; NID:g3261720; PIDN:CAB07046.1; PID:e306646;
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: RV0140

Query Match 2.1%; Score 7; DB 2; Length 126;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 RIVLEPS 71
 |||||
 DB 4 RIVLEPS 10

RESULT 8

hypothetical protein PH0659 - *Pyrococcus horikoshii*
 C:Species: *Pyrococcus horikoshii*
 C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 05-Nov-1999
 C:Accession: D71111
 R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin, M.; Ohtuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, DNA Res. 5, 55-76, 1998
 A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
 A:Reference number: A71000; MUID:98344137; PMID:9679194
 A:Accession: D71111
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-158 <RAM>
 A:Cross-references: GB:AP000003; NID:g3236130; PIDN:BA29750.1; PID:dl030693; PID:g32570
 A:Experimental source: strain OF3
 A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
 C:Genetics:
 A:Gene: PH0659

Query Match 2.1%; Score 7; DB 2; Length 158;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 ALFSAFL 216
 |||||
 DB 66 ALFSAFL 72

RESULT 9

S45632
 H+-transporting two-sector ATPase (EC 3.6.3.14) epsilon chain precursor - yeast (*Saccharomyces cerevisiae*)
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 03-May-1994 #sequence_revision 27-Jan-1995 #text_change 03-Jun-2002
 C:Accession: S45632; S47658; S50978; S52515; S67536; S39208
 R:Giraud, M.F.; Velours, J.; Eur. J. Biochem. 222, 851-859, 1994
 A:Title: ATP synthase of yeast mitochondria. Isolation of the F(1)-delta subunit, seq
 A:Reference number: S45632; MUID:94298825; PMID:8026496
 A:Accession: S45632

A:Molecule type: DNA
 A:Residues: 1-160 <GIR>
 A:Cross-references: EMBL:Z21857; NID:9433619; PIDN:CAA79912.1; PID:9433620
 A:Accession: S47658
 A:Molecule type: protein
 A:Residues: 23-72,112-121,141-160 <GIR>
 R:Murphy, L.; Richards, C.; Gentles, S.; Harris, D.
 A:Title: ATP synthase of yeast mitochondria. Isolation of the F(1)-delta subunit, seq
 A:Reference number: S50976
 A:Accession: S50978
 A:Molecule type: DNA
 A:Residues: 1-160 <MGR>
 A:Cross-references: EMBL:Z48008; NID:9642799; PIDN:CAA88057.1; PID:9642802
 R:Andre, B.; Vissers, S.; Urrestarazu, L.
 A:Title: The sequence of a 42 kb segment located on the left arm of chromosome
 A:Reference number: S52492
 A:Accession: S52515

A:Molecule type: DNA
 A:Residues: 1-160 <AND>
 A:Cross-references: EMBL:Z48432; NID:9683669; PIDN:CAA88355.1; PID:9683693
 R:Urrestarazu, L.A.; Andre, B.; Vissers, S.
 A:Title: The sequence of a 42 kb segment located on the left arm of chromosome
 A:Reference number: S67535
 A:Accession: S67536
 A:Molecule type: DNA
 A:Residues: 1-160 <URR>
 A:Cross-references: EMBL:Z74052; NID:91430960; PIDN:CAA98560.1; PID:91430961; MIPS:YD
 A:Experimental source: strain S288C
 C:Genetics:
 A:Gene: SGD:ATP16; ATP14
 A:Cross-references: MIPS:YDL004W; SGD:S0002162
 A:Map position: 4L
 A:Genome: nuclear
 A:Note: YDL004W

C:Superfamily: H+-transporting ATP synthase epsilon chain
 C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrial
 F:1-22/Domain: trans peptide (mitochondrion) #status predicted <TMP>
 F:23-160/Product: H+-transporting ATP synthase epsilon chain #status predicted <MAT>

Query Match 2.1%; Score 7; DB 2; Length 160;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 309 RSIIGKS 315
 |||||
 DB 3 RSIIGKS 9

RESULT 10

coat protein - tobacco necrosis satellite virus

C:Species: tobacco necrosis satellite virus
C:Date: 31-Mar-1981 #sequence_revision 31-Mar-1981 #text_change 23-Jul-1999
C:Accession: A04217
R:Ysebaert, M.; van Emmelo, J.; Fiers, W.
J. Mol. Biol. 143, 273-287, 1980
A:Title: Total nucleotide sequence of a nearly full-size DNA copy of satellite tobacco r
A:Reference number: A04217; MUID:81170584; PMID:6260960
A:Accession: A04217
A:Molecule type: genomic RNA
A:Residues: 1-196 <YSE>
A:Cross-references: GB:J02399; NID:g335254; PIDN:AAA69583.1; PID:g335255
C:Superfamily: satellite tobacco necrosis virus coat protein
C:Keywords: coat protein

Query Match 2.1%; Score 7; DB 1; Length 196;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 184 LTGSIR 190
138 LTGSIR 144

RESULT 11
C70909
Probable two-component sensor - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: C70909
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garler, T.; Churcher, C.; Harris, D.; Gordon, S.
Rajandream, M.A.; Rogers, J.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.
Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:9825987; PMID:9634230
A:Accession: C70909
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-207 <COL>
A:Cross-references: GB:297182; GB:AL123456; NID:g3250720; PIDN:CAB09983.1; PID:el192556;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV0600c

Query Match 2.1%; Score 7; DB 2; Length 207;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 57 SALETR 63
7 SALETR 13

RESULT 12
T18109
anryin repeat protein A607R - Chlorella virus PBCV-1
C:Species: Chlorella virus PBCV-1
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18109
R:Graves, M.V.; Van Etten, J.L.
submitted to the EMBL Data Library, May 1999
A:Reference number: Z18806
A:Accession: T18109
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-229 <GRA>
A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96943.1
A:Experimental source: specific host Chlorella strain NC64A
C:Genetics:
A:Note: A607R

Query Match 2.1%; Score 7; DB 2; Length 229;

Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 272 KELRDF 278
16 KELRDF 22

RESULT 13
A10267
Probable membrane protein YPO2199 [Imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: A10267
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: A10267
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-256 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC91005.1; PID:g15980199; GSPDB:GN00175
C:Genetics:
A:Gene: YPO2199
C:Superfamily: ypic protein

Query Match 2.1%; Score 7; DB 2; Length 256;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 43 LRVSAV 49
81 LRVSAV 87

RESULT 14
T34691
hypothetical protein SC1B5.13c SC1B5.13c - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 21-Jan-2000
C:Accession: T34691
R:Harris, D.; Taylor, K.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, May 1998
A:Reference number: Z21553
A:Accession: T34691
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-262 <HAR>
A:Cross-references: EMBL:AL023517; PIDN:CAA19987.1; GSPDB:GN00070; SCODEB:SC1B5.13c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCODEB:SC1B5.13c
C:Superfamily: erythrocyte band 7 integral membrane protein

Query Match 2.1%; Score 7; DB 2; Length 262;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 309 RSIIKS 315
121 RSIIKS 127

RESULT 15
C70478
proteinase IV - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 08-Oct-1999
C:Accession: C70478
R:Decker, G.; Warren, P.V.; Gasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;

V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: C70478
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-283 <AOE>
A:Cross-references: GB:AE000771; NID:g2984286; PIDN:AA07815.1; PID:g2984291; GB:AE00065
A:Experimental source: strain VFS
C:Genetics:
A:Gene: sppA

	Query Match	2.1%	Score 7;	DB 2;	Length 283;
	Best Local Similarity	100.0%	Pred. No. 42;		
	Matches 7;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	261 LKRLVG 267				
Db	256 LKRLVG 262				

Search completed: June 24, 2003, 18:56:50
job time : 17 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 18:50:11 ; Search time 16 Seconds
(without alignments)
852.857 Million cell updates/sec

Title: US-09-870-406A-33

Sequence: 1 MALSMERGFSGSCFKAPNP.....SIIGKSYETRPMDLTGQFIG 329

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size: 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	2.7	477	1	MYPH_HUMAN
2	8	2.4	944	1	DNL4_YEAST
3	7	2.1	59	1	MTBC_METTE
4	7	2.1	120	1	RL17_BACHD
5	7	2.1	160	1	ATPD_YEAST
6	7	2.1	196	1	COAT_STN1
7	7	2.1	212	1	MTBC_METBA
8	7	2.1	213	1	MTBC_METAC
9	7	2.1	213	1	MTBC_METAC
10	7	2.1	214	1	MTBC_METMA
11	7	2.1	216	1	MTBC_METMA
12	7	2.1	216	1	MTBC_METAC
13	7	2.1	277	1	DLHD_RHET
14	7	2.1	289	1	YHDF_BACSU
15	7	2.1	324	1	PSC2_MYCTU
16	7	2.1	330	1	UL16_HSVSA
17	7	2.1	344	1	ISPG_THEMA
18	7	2.1	345	1	AFOR_MOUSE
19	7	2.1	356	1	AFOR_MOUSE
20	7	2.1	387	1	PKK_PSEAF
21	7	2.1	392	1	FTZ2_ARCFU
22	7	2.1	465	1	GAR2_HUMAN
23	7	2.1	465	1	GAR2_HUMAN
24	7	2.1	465	1	GAR2_MOUSE
25	7	2.1	473	1	GAR2_MOUSE
26	7	2.1	474	1	GAR1_MOUSE
27	7	2.1	474	1	GAR1_MOUSE
28	7	2.1	477	1	DLHD_PSEFL
29	7	2.1	583	1	FOJO_DROME
30	7	2.1	683	1	TKT1_ECOLI
31	7	2.1	692	1	TKH3_YEAST
32	7	2.1	824	1	JIP2_HUMAN
33	7	2.1	830	1	JIP2_MOUSE

34	7	2.1	941	1	CHRD_XENLA	091713 xenopus lae
35	7	2.1	1350	1	VG72_HSV1	000103 ictaluriid h
36	6	1.8	57	1	V556_ARCFU	028816 archaeoglob
37	6	1.8	62	1	V452_BPR69	064300 bacteriophob
38	6	1.8	64	1	RK35_CYACA	094119 cyanidium c
39	6	1.8	76	1	CD24_MOUSE	P24807 mus musculu
40	6	1.8	81	1	EX7S_PASMU	09cna0 pasteurella
41	6	1.8	82	1	VRDX_PLEBO	P46043 plectonema
42	6	1.8	84	1	DHSD_CHOCR	P54323 chondrus cr
43	6	1.8	114	1	Y941_ARCFU	029321 archaeoglob
44	6	1.8	117	1	CHH3_PENP	094676 penaeus jap
45	6	1.8	117	1	RL20_CAMJE	09p1q0 campylobact

ALIGNMENTS

RESULT 1
ID MYPH_HUMAN STANDARD: PRT: 477 AA.
AC Q13203; Q16886;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin-binding protein H (MyBP-H) (H-protein).
GN MYBPH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=9352409; PubMed=8486381;
RA Vaughan K.T., Weber F.E., Ried T., Ward D.C., Rejnach F.C.,
RT Fischenman D.A.;
RT "Human myosin-binding protein H (MyBP-H): complete primary sequence,
RT genomic organization, and chromosomal localization.";
RL Genomics 16:34-40(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Whittle M.R., Fischenman D.A., Rejnach F.C.;
RL Submitted (MAY-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: BINDS TO MYOSIN; PROBABLY INVOLVED IN INTERACTION WITH
CC THICK MYOFILAMENTS IN THE A-BAND.
CC -1- TISSUE SPECIFICITY: SKELETAL MUSCLE.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. MYBP
CC SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; L05606; AAA6339.1; -;
CC EMBL; L05607; AAA02904.2; -;
CC EMBL; U27266; AAB86737.1; -;
CC HSPB; P56276; ITLK.
CC GeneW; HGNC:7552; MYBPH.
CC MIM; 160795; -;
CC InterPro: IPR003961; FN.III.
CC InterPro: IPR003962; FN.II.Repeat.
CC InterPro: IPR003006; IG.MHC.
CC InterPro: IPR003598; IG.C2.
CC InterPro: IPR003600; IG.Like.
CC Pfam; PF00041; In3; 2.
CC Pfam; PF00047; Ig; 2.
CC PRINTS; PR00014; FNTYPEIII.

DR SMART; SM00060; FN3; 2.
 DR SMART; SM00410; IG-like; 1.
 DR SMART; SM00408; IGC2; 1.
 KW Immunoglobulin domain; Cell adhesion; Muscle protein; Thick filament;
 KM Repeat.
 FT DOMAIN 71 157 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 189 248 IG-LIKE C2-TYPE DOMAIN 1.
 FT DOMAIN 267 352 FIBRONECTIN TYPE-III 2.
 FT DOMAIN 396 456 IG-LIKE C2-TYPE DOMAIN 2.
 FT CONFLICT 114 114 A -> G (IN REF. 2).
 FT CONFLICT 149 150 SS -> RP (IN REF. 2).
 FT CONFLICT 271 272 SS -> CC (IN REF. 1).
 FT CONFLICT 406 411 RASPKP -> ELHQA (IN REF. 1).
 SQ SEQUENCE 477 AA; 52045 MW; F27ABE8ABCFE77F CRC64;

Query Match 2.7%; Score 9; DB 1; Length 477;
 Best Local Similarity 100.0%; Pred. No. 0.26;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 42 FILRVSAYS 50
 Db 141 FILRVSAYS 149

RESULT 2
 DN14_YEAST STANDARD; PRT; 944 AA.

AC 008387; 002914; 002913;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA ligase II (EC 6.5.1.1) (Polydeoxyribonucleotide synthase [ATP])
 DE (DNA ligase IV homolog).
 GN DN14 OR YOR005C OR UND407 OR UNE452.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97051599; PubMed=8896276;
 RA Sterky F., Holmberg A., Pettersson B., Uhlen M.;
 RT "The sequence of a 30 kb fragment on the left arm of chromosome XV
 from Saccharomyces cerevisiae reveals 15 open reading frames, five of
 which correspond to previously identified genes.";
 RL Yeast 12:1091-1095(1996).
 RN [2]
 RP REVISIONS.
 RA Pettersson B., Sterky F., Uhlen M.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=97248602; PubMed=9092653;
 RA Ramos W., Tappe N., Talamantez J., Friedberg E.C., Tomlinson A.E.;
 RT "Two distinct DNA ligase activities in mitotic extracts of the yeast
 Saccharomyces cerevisiae.";
 RL Nucleic Acids Res. 25:1485-1492(1997).
 CC -1- FUNCTION: HAS MINOR DNA JOINING ACTIVITY. CAN ACT ON
 OLIGO(PdT)/POLY(RA) SUBSTRATE.
 CC -1- CATALYTIC ACTIVITY: ATP + (deoxyribonucleotide)(N) +
 (deoxyribonucleotide)(M) -> AMP + diphosphate +
 (deoxyribonucleotide)(N+M).
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT DNA LIGASE FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 BCT DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; Z74913; CAA9193.1; -
 DR EMBL; U43491; AAC49485.1; ALT_SEQ.
 DR EMBL; U43491; AAC49484.1; ALT_SEQ.
 DR SCD; S0005531; DN14.
 DR InterPro; IPR001357; BRC1.
 DR InterPro; IPR000977; DNA_Ligase.
 DR Pfam; PF00533; BRC1; 2.
 DR Pfam; PF01068; DNA_Ligase; 1.
 DR SMART; SM00292; BRC1; 1.
 DR TIGRfam; TIGR00574; dnl1; 1.
 DR PROSITE; PS0172; BRC1; 2.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 DR PROSITE; PS00333; DNA_LIGASE_A2; 1.
 DR PROSITE; PS0160; DNA_LIGASE_A3; 1.
 KW DNA repair; DNA replication; DNA recombination; Cell division; Ligase;
 KM ATP-binding; Nuclear protein; Repeat.
 FT DOMAIN 681 780 BRC1 1.
 FT BINDING 840 941 BRC1 2.
 FT BINDING 282 282 AMP (BY SIMILARITY).
 SQ SEQUENCE 944 AA; 108514 MW; 7FB6D7927E1719B5 CRC64;

Query Match 2.4%; Score 8; DB 1; Length 944;
 Best Local Similarity 100.0%; Pred. No. 5.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 186 GESIKFFS 193
 Db 296 GESIKFFS 303

RESULT 3

MTBC_METTE STANDARD; PRT; 59 AA.

AC Q9YGA3;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Dimethylamine corrinoid protein (Fragment).
 GN MTBC.
 OS Methanosarcina thermophila.
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 1825 / TM-1;
 RX MEDLINE=20225851; PubMed=10762254;
 RA Paul L., Ferguson D.J., Krzycki J.A.;
 RT "The trimethylamine methyltransferase gene and multiple dimethylamine
 methyltransferase genes of Methanosarcina barkeri contain in-frame
 RT and read-through amber codons.";
 RL J. Bacteriol. 182:2520-2529(2000).
 CC -1- FUNCTION: Acts as a methyl group carrier between mtbB and mtbA (By
 similarity).
 CC -1- PATHWAY: Methanogenesis from dimethylamines.
 CC -1- SIMILARITY: BELONGS TO THE METHYLAMINE CORRINOID PROTEIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; AF153452; AAD38788.1; -
 DR InterPro; IPR003312; B12-binding.
 DR Pfam; PF02310; B12-binding; 1.
 KW Cobalt; Methanogenesis.
 FT NON_TER 1 1
 SQ SEQUENCE 59 AA; 6257 MW; 479E0F91115B5EFC CRC64;

```

Query Match      2.1%; Score 7; DB 1; Length 59;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      269 AKAKEL 275
DB      52 AKAKEL 58

RESULT 4
RL17_BACHD      STANDARD;      PRT;      120 AA.
ID      RL17_BACHD
AC      O50635; O9MMJ4; O9JFW2;
DT      15-DEC-1998 (Rel. 37, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      50S ribosomal protein L17.
GN      RPLQ OR BH0163.
OS      Bacillus halodurans.
OC      Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX      NCBI_TaxID=86665;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C-125 / JCM 9153;
RX      MEDLINE=99209008; PubMed=10192928;
RA      Takami H., Takaki Y., Nakasone K., Hirama C., Inoue A., Horikoshi K.;
RT      "Sequence analysis of a 32-kb region including the major ribosomal
RT      protein gene clusters from alkaliphilic Bacillus sp. strain C-125.";
RL      Biosci. Biotechnol. Biochem. 63:452-455(1999).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C-125 / JCM 9153;
RX      MEDLINE=20512582; PubMed=11058132;
RA      Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA      Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA      Horikoshi K.;
RT      "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT      halodurans and genomic sequence comparison with Bacillus subtilis.";
RL      Nucleic Acids Res. 28:4317-4331(2000).
RN      [3]
RP      SEQUENCE OF 1-83 FROM N.A.
RC      STRAIN=C-125 / JCM 9153;
RX      MEDLINE=99052103; PubMed=9835038;
RA      Nakasone K., Takaki Y., Takami H., Inoue A., Horikoshi K.;
RT      "Cloning and expression of the gene encoding RNA polymerase alpha
RT      subunit from alkaliphilic Bacillus sp. strain C-125.";
RL      FEBS Microbiol. Lett. 168:269-276(1998).
CC      -1- SIMILARITY: BELONGS TO THE L17P FAMILY OF RIBOSOMAL PROTEINS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AB017508; BAA75299.1; -
DR      EMBL; AF001507; BAB03862.1; -
DR      EMBL; AB010082; BAA24195.1; -
DR      InterPro; IPR000456; Ribosomal_L17.
DR      Pfam; PF01196; Ribosomal_L17; 1.
DR      ProDom; PD004277; Ribosomal_L17; 1.
DR      TIGRfam; TIGR00059; L17; 1.
DR      PROSITE; PS01167; RIBOSOMAL_L17; 1.
KW      Ribosomal protein; Complete proteome.
SQ      SEQUENCE 120 AA; 13393 MW; DB5C9F1F26515C35 CRC64;

Query Match      2.1%; Score 7; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      268 EAKAKEL 274

```

```

DB      34 EAKAKEL 40

RESULT 5
ATPD_YEAST
ID      ATPD_YEAST      STANDARD;      PRT;      160 AA.
AC      012165;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      ATP synthase delta chain, mitochondrial precursor (EC 3.6.3.14).
GN      ATP16 OR YDL004W OR YDL119.03 OR D2935.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX      NCBI_TaxID=4932;
RN      [1]
RP      SEQUENCE FROM N.A. AND SEQUENCE OF 23-72; 112-121 AND 141-160.
RC      STRAIN=D273-10B/4;
RX      MEDLINE=94298825; PubMed=8026496;
RA      Giraud M.-F., Velours J.;
RT      "ATP synthase of yeast mitochondria. Isolation of the F1 delta
RT      subunit, sequence and disruption of the structural gene.";
RL      Eur. J. Biochem. 222:851-859(1994).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=S288c / AB972;
RA      Murphy L., Richards C., Gentles S., Harris D., Barrell B.G.,
RA      Rajandream M.A.;
RL      Submitted (JAN-1995) to the EMBL/Genbank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=S288c;
RA      Andre B., Vissers S., Urrestarazu L.;
RT      submitted (FEB-1995) to the EMBL/Genbank/DBJ databases.
CC      -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC      GRADIENT ACROSS THE MEMBRANE. MAY BE INVOLVED IN THE STABILITY
CC      OF THE CATALYTIC SECTOR.
CC      -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(in) - ADP + phosphate +
CC      H(+)(out).
CC      -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC      CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC      SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC      HAS THREE MAIN SUBUNITS: A, B AND C.
CC      -1- SUBCELLULAR LOCATION: Mitochondrial.
CC      -1- SIMILARITY: BELONGS TO THE ATPASE EPSILON CHAIN FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; Z21857; CAA79912.1; -
DR      EMBL; Z48008; CAA88057.1; -
DR      EMBL; Z48432; CAA88355.1; -
DR      EMBL; Z74052; CAA98560.1; -
DR      SGD; S0002162; ATP16.
DR      InterPro; IPR001469; ATPsynth_DE.
DR      Pfam; PF02823; ATP-synth_DE_N; 1.
DR      ProDom; PD000944; ATPsynth_DE; 1.
DR      TIGRfam; TIGR01216; ATP_synth_eps1; 1.
KW      ATP synthetase; CF(1); Hydrogen ion transport; Hydrolase;
KW      Mitochondrion; Transit peptide.
FT      TRANSIT 1
FT      CHAIN 23 160 MITOCHONDRION.
SQ      SEQUENCE 160 AA; 17020 MW; AEBDDA364107977 CRC64;

Query Match      2.1%; Score 7; DB 1; Length 160;
Best Local Similarity 100.0%; Pred. No. 12;

```



```

ID  MB01_METAC  STANDARD;  PRT;  213 AA.
AC  Q8RTB0;
DT  15-JUN-2002 (Rel. 41, Created)
DT  15-JUN-2002 (Rel. 41, Last sequence update)
DE  15-JUN-2002 (Rel. 41, Last annotation update)
GN  Dimethylamine corrinoid protein 1.
OS  MB01 OR MA0527.
OC  Methanosarcina acetivorans.
OC  Archaea: Euryarchaeota: Methanococci: Methanosarcinales;
OX  Methanosarcinaceae; Methanosarcina.
OX  NCBI_TaxID=2214;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=C2A / ATCC 35395 / DSM 2834;
RX  MEDLINE=21929760; PubMed=11932238;
RA  Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA  Fitzhugh W., Calvo S., Engels R., Smirnov S., Alnoor D., Brown A.,
RA  Allen N., Naylor J., Stange-Thomann N., Deatellano K., Johnson R.,
RA  Linton L., McEwan P., McKernan K., Talamas J., Tjirell A., Ye W.,
RA  Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA  Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA  Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA  Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA  Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA  Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA  Metcalf W.W., Birren B.;
RT  "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT  and physiological diversity.";
RL  Genome Res. 12:532-542(2002).
CC  -1- FUNCTION: Acts as a methyl group carrier between mtbb and mtba (By
CC  similarity).
CC  -1- PATHWAY: Methanogenesis from dimethylamines.
CC  -1- SIMILARITY: BELONGS TO THE METHYLAMINE CORRINOID PROTEIN FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@sib-sib.ch).
CC  -----
CC  EMBL: AE010713; AAM03971.1; -
CC  DR  Cobalt; Methanogenesis; Complete proteome.
CC  KW  SEQUENCE 213 AA; 22436 MW; 8A0AF2A9C1D8C8E4 CRC64;
SQ
Query Match 2.1%; Score 7; DB 1; Length 213;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 269 AKARELL 275
DB 206 AKARELL 212

```

```

RA  Fitzhugh W., Calvo S., Engels R., Smirnov S., Alnoor D., Brown A.,
RA  Allen N., Naylor J., Stange-Thomann N., Deatellano K., Johnson R.,
RA  Linton L., McEwan P., McKernan K., Talamas J., Tjirell A., Ye W.,
RA  Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA  Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA  Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA  Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA  Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA  Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA  Metcalf W.W., Birren B.;
RT  "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT  and physiological diversity.";
RL  Genome Res. 12:532-542(2002).
CC  -1- FUNCTION: Acts as a methyl group carrier between mtbb and mtba (By
CC  similarity).
CC  -1- PATHWAY: Methanogenesis from dimethylamines.
CC  -1- SIMILARITY: BELONGS TO THE METHYLAMINE CORRINOID PROTEIN FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@sib-sib.ch).
CC  -----
CC  EMBL: AE010934; AAM05810.1; -
CC  DR  Cobalt; Methanogenesis; Complete proteome.
CC  KW  SEQUENCE 213 AA; 22720 MW; 5BAF9326053D06B9 CRC64;
SQ
Query Match 2.1%; Score 7; DB 1; Length 213;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 269 AKARELL 275
DB 206 AKARELL 212

```

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE013550; AAM32657.1; ALT_INT.
 KW Cobalt; Methanogenesis.
 SQ SEQUENCE 214 AA; 22674 MW; 2E3CA138C0A44DAF CRC64;
 Query Match
 Best Local Similarity 2.1%; Score 7; DB 1; Length 214;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 269 ARAKELL 275
 DB 207 ARAKELL 213
 RESULT 11
 MBL1_METMA
 ID MBL1_METMA STANDARD; PRT; 216 AA.
 AC P58979;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Dimethylamine corrinoid protein 1.
 GN MBL1 OR M2052.
 OS Methanosarcina mazei (Methanosarcina frisia).
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2209;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Goel / GGT / ATCC BAA-199 / DSM 3647 / OCM 88;
 RX MEDLINE=22120827; PubMed=12125824;
 RA Deppeleier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
 RA Martinez-Arias R., Henne A., Wleker A., Baesner S., Jacobl C.,
 RA Bruggemann H., Lienard T., Christmann A., Boemcke M., Steckel S.,
 RA Bhattacharya A., Lykdis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
 RA Fritz H.-J., Gottschalk G.;
 RT "The genome of Methanosarcina mazei: evidence for lateral gene
 transfer between Bacteria and Archaea."
 RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
 CC -1- FUNCTION: Acts as a methyl group carrier between mtbB and mtbA (by
 similarity).
 CC -1- PATHWAY: Methanogenesis from dimethylamines.
 CC -1- SIMILARITY: BELONGS TO THE METHYLAMINE CORRINOID PROTEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE013445; AAM31748.1; ALT_INT.
 KW Cobalt; Methanogenesis.
 SQ SEQUENCE 216 AA; 22980 MW; B8B76B04651FA76 CRC64;
 Query Match
 Best Local Similarity 2.1%; Score 7; DB 1; Length 216;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 269 ARAKELL 275
 DB 207 ARAKELL 213
 RESULT 12
 MBL2_METAC
 ID MBL2_METAC STANDARD; PRT; 216 AA.
 AC O8T571;
 DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Dimethylamine corrinoid protein 2.
 GN MBL2 OR MA0934.
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2214;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238;
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Alnoor D., Brown A.,
 RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
 RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kuettner H.C., Kizycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
 and physiological diversity."
 RL Genome Res. 12:532-542(2002).
 CC -1- FUNCTION: Acts as a methyl group carrier between mtbB and mtbA (by
 similarity).
 CC -1- PATHWAY: Methanogenesis from dimethylamines.
 CC -1- SIMILARITY: BELONGS TO THE METHYLAMINE CORRINOID PROTEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE010758; AAM04367.1; ALT_INT.
 KW Cobalt; Methanogenesis; Complete proteome.
 SQ SEQUENCE 216 AA; 23101 MW; F18A97039F4B57FA CRC64;
 Query Match
 Best Local Similarity 2.1%; Score 7; DB 1; Length 216;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 269 ARAKELL 275
 DB 207 ARAKELL 213
 RESULT 13
 DLDH_RHIEF
 ID DLDH_RHIEF STANDARD; PRT; 277 AA.
 AC O05940;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Probable dihydrolipoamide dehydrogenase (EC 1.8.1.4) (E3 component of
 DE pyruvate and 2-oxoglutarate dehydrogenases complexes) (ORF-E3)
 DE (Fragment).
 DE Rhizobium etl.
 OS Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Bacteria; Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=29449;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CE3;
 RX MEDLINE=97286510; PubMed=9141657;
 RA Tate R., Riccio A., Iaccarino M., Patriarca E.J.;
 RT "Cloning and transcriptional analysis of the lipA (lipoic acid
 synthetase) gene from Rhizobium etl.";

RL EMBL Microbiol. Lett. 149:165-172(1997).
 CC -1- FUNCTION: LIPAMIDE DEHYDROGENASE IS A COMPONENT OF THE ALPHA-
 CC KETOACID DEHYDROGENASE COMPLEXES (BY SIMILARITY).
 CC CATALYTIC ACTIVITY: Dihydrolipoamide + NAD(+) = Lipoamide + NADH.
 CC -1- COFACTOR: FAD (BY SIMILARITY).
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -1- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE
 CC OXIDOREDUCTASES CLASS-I.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Y11708; CAA72399.1; -.
 DR HSSP: P18925; 3LAD.
 DR InterPro: IPR001327; FAD_Pyr_redox.
 DR InterPro: IPR001100; Pyr_redox.
 DR InterPro: IPR004099; Pyr_redox_dim.
 DR Pfam: PF00070; Pyr_redox; 1.
 DR Pfam: PF02852; Pyr_redox_dim; 1.
 DR PRINTS: PR00411; PNDPDMASE1.
 DR ProDom: PD000139; FAD_Pyr_redox; 1.
 DR PROSITE: PS00076; PYRIDINE_REDOX_1; PARTIAL.
 DR Redox-active center: Glycolysis; Oxidoreductase; NAD; Flavoprotein;
 DR FAD.
 FM NON_TER 1 1
 FT NP_BIND 115 125 FAD (FLAVIN PART) (BY SIMILARITY).
 FT ACT_SITE 259 259 BY SIMILARITY.
 SQ SEQUENCE 277 AA; 29547 MW; CD3E35DB1FE4CAF CRC64;

 Query Match 2.1%; Score 7; DB 1; Length 277;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 268 EAKAKEL 274
 Db 178 EAKAKEL 184

 RESULT 14
 YHDF_BACSU STANDARD; PRT; 289 AA.
 ID YHDF_BACSU
 AC 007575;
 DT 15-JUN-1998 (Rel. 36, Created)
 DT 15-JUN-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical oxidoreductase yndf (EC 1.-.-.-).
 GN YHDF.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codant J.J., Conneton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehlich S.D., Emmerson P.T.,
 RA Entlian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,

RA Filtz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Haeche J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koelter P., Konigstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Nodack M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudga B., Park S.H.,
 RA Paro V., Pohl T.M., Portetelle D., Potwolik S., Prescott A.M.,
 RA Presseau E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Yamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tesato V., Uchiyama S., Vandenbol M., Vannier F., Vassartoli A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Welternegger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
 RA "the complete genome sequence of the Gram-positive bacterium Bacillus
 RA subtilis."
 RL Nature 390:249-256(1997).
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC (SDR) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Y14082; CAA74490.1; -.
 DR EMBL: Z99109; CAB12784.1; -.
 DR HSSP: Q12634; 1YBV.
 DR Subtilist: BG13012; yndf.
 DR InterPro: IPR002198; ADH_short.
 DR Pfam: PF00106; adh_short; 1.
 DR PROSITE: PS00061; ADH_SHORT; 1.
 FM NP_BIND 49 73
 FT ACT_SITE 194 194 NAD OR NADP (BY SIMILARITY).
 SQ SEQUENCE 289 AA; 31509 MW; 4EC98C62B9241B01 CRC64;

 Query Match 2.1%; Score 7; DB 1; Length 289;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 60 EETRRRI 66
 Db 84 EETRRRI 90

 RESULT 15
 PSC2_MYCTU STANDARD; PRT; 324 AA.
 ID PSC2_MYCTU
 AC 086344; Q50797;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Phosphate transport system permease protein pstC-2.
 GN PSTC2 OR RV0929 OR MT0956 OR MYCT21C12.23.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Erdmann;
 RX MEDLINE=97000022; PubMed=8843165;

Job time : 17 secs

```

RA Bralbant M., Lefevre P., de Wit L., Ooms J., Peirs P., Huygen K.,
RA Watiez R., Content J.;
RT "Identification of a second Mycobacterium tuberculosis gene cluster
RT encoding proteins of an ABC phosphate transporter.";
RL FEBS Lett. 394:206-212(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-H37RV;
RX MEDLINE-98295987; PubMed-9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouiri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
CC FOR PHOSPHATE. PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF THE
CC SUBSTRATE ACROSS THE MEMBRANE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -1- SIMILARITY: BELONGS TO THE BINDING-PROTEIN-DEPENDENT TRANSPORT
CC SYSTEM PERMEASE FAMILY. CYSTW SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z47983; CA88026.1; -
DR EMBL: Z95209; CAB08489.1; -
DR EMBL: AE006981; AAK45203.1; -
DR Tuberculist; RV0929; -
DR InterPro: IPR000515; BPD.transp.
DR Pfam: PF00528; BPD.transp. 1.
DR PROSITE: PS00402; BPD_TRANSP_INN_MEMBR. 1.
KW Transport; Phosphate transport; Transmembrane; Complete proteome.
FT TRANSMEM 28 50 POTENTIAL.
FT TRANSMEM 87 109 POTENTIAL.
FT TRANSMEM 122 144 POTENTIAL.
FT TRANSMEM 172 194 POTENTIAL.
FT TRANSMEM 235 257 POTENTIAL.
FT TRANSMEM 291 313 POTENTIAL.
FT TRANSMEM 319 342 MISSING (IN REF. 1).
SQ SEQUENCE 324 AA; 34261 MW; 0C54657C8A3CEAC5 CRC64;

```

Query Match 2.1%; Score 7; DB 1; Length 324;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 132 ANFFTST 138
 |||||
 Db 61 ANFFTST 67

Search completed: June 24, 2003, 18:54:57

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 12:30:29 ; Search time 80 Seconds
(Without alignments)
8280.273 Million cell updates/sec

Title: US-09-870-406A-32

Perfect score: 2160

Sequence: 1 gaattcccccagtcacgtg.....tcgtctctactaatacaca 2160

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCNUS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/Backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	53.6	2.5	7218	1 US-08-232-463-14	Sequence 14, Appl
C 2	43.2	2.0	7218	1 US-08-232-463-14	Sequence 14, Appl
C 3	40.2	1.9	5361	4 US-08-973-462-2	Sequence 2, Appl
C 4	40.2	1.9	6152	4 US-08-973-462-1	Sequence 13, Appl
C 5	39.4	1.8	19124	2 US-08-487-8268-13	Sequence 3, Appl
C 6	38.6	1.8	12132	4 US-09-741-150-3	Sequence 11, Appl
C 7	38.2	1.8	2251	4 US-08-991-677-11	Sequence 14, Appl
C 8	38	1.8	9472	1 US-08-325-547-9	Sequence 9, Appl
C 9	37.2	1.7	522	4 US-09-134-001C-1438	Sequence 14, Appl
C 10	36.4	1.7	2556	4 US-08-811-682-18	Sequence 14, Appl
C 11	36.4	1.7	3225	1 US-08-487-090-14	Sequence 14, Appl
C 12	36.4	1.7	3225	2 US-08-487-700B-7	Sequence 14, Appl
C 13	36.4	1.7	3225	2 US-09-007-383-14	Sequence 14, Appl
C 14	36.4	1.7	9471	4 US-08-811-682-1	Sequence 1, Appl
C 15	36.2	1.7	9636	1 US-08-323-170B-1	Sequence 1, Appl
C 16	36.2	1.7	9636	1 US-08-954-441-1	Sequence 1, Appl
C 17	35.4	1.6	289	4 US-09-007-005-17	Sequence 17, Appl
C 18	35.4	1.6	289	4 US-09-244-796-17	Sequence 17, Appl
C 19	35	1.6	5852	1 US-07-867-106-2	Sequence 1, Appl
C 20	34.8	1.6	2760	4 US-09-198-484-1	Sequence 2, Appl
C 21	34.8	1.6	6265	4 US-09-129-112-3	Sequence 3, Appl
C 22	34.8	1.6	50000	4 US-09-146-053-4	Sequence 3, Appl
C 23	34.6	1.6	162450	4 US-09-345-882-1	Sequence 1, Appl
C 24	34.4	1.6	432	4 US-09-008-687A-19	Sequence 19, Appl
C 25	34.4	1.6	2272	4 US-09-453-702B-126	Sequence 126, Ap
C 26	34.2	1.6	432	4 US-09-134-001C-1126	Sequence 1126, Ap
C 27	34	1.6	2098	2 US-08-937-931-5	Sequence 5, Appl

C 28	34	1.6	2098	4 US-09-285-502-5	Sequence 5, Appl
C 29	34	1.6	2098	4 US-09-709-126-5	Sequence 5, Appl
C 30	34	1.6	2098	4 US-09-871-385A-5	Sequence 5, Appl
C 31	33.8	1.6	462	4 US-08-964-652-3	Sequence 3, Appl
C 32	33.8	1.6	1236	4 US-08-964-652-1	Sequence 1, Appl
C 33	33.6	1.6	785	4 US-09-276-141-11	Sequence 11, Appl
C 34	33.6	1.6	6243	2 US-09-056-075-1	Sequence 1, Appl
C 35	33.2	1.5	2789	4 US-08-961-527-254	Sequence 254, App
C 36	33.2	1.5	7493	1 US-08-212-133A-7	Sequence 7, Appl
C 37	33.2	1.5	7493	3 US-08-474-503-5	Sequence 5, Appl
C 38	33.2	1.5	7493	2 US-08-670-707A-5	Sequence 5, Appl
C 39	33.2	1.5	7493	4 US-09-037-601-5	Sequence 5, Appl
C 40	33.2	1.5	7493	4 US-09-315-179-5	Sequence 5, Appl
C 41	33.2	1.5	7493	5 PCT-US94-13200-5	Sequence 5, Appl
C 42	33	1.5	1755	4 US-08-543-246B-8	Sequence 8, Appl
C 43	33	1.5	3620	4 US-09-446-504-55	Sequence 55, Appl
C 44	33	1.5	3620	4 US-09-712-266-55	Sequence 55, Appl
C 45	33	1.5	5176	4 US-09-610-040-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 3670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)636-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZapT-F15
US-08-232-463-14
Query Match 2.5%; Score 53.6; DB 1; Length 7218;

Best Local Similarity 2.1%; Pred. No. 6.9e-05;
Matches 8; Conservative 223; Mismatches 147; Indels 0; Gaps 0;

```

QY 1633 AAAAGATAGTAGTGAAGCAAGCAAGATTAAGATTTGATCCATTTAGTGTCC 1692
Db 1447 AAGATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1388
QY 1693 CATTATTAATTAAGCTTGAAGATGTGAAGATTTGACAAATTCAGAGCTGTAA 1752
Db 1387 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1328
QY 1753 GGGATTTCTGTTCAATGGGGTGTAGTGTAGCAACAAATCATTTGATTTCTTC 1812
Db 1327 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1268
QY 1813 CAGATGACCAAGCAAGATGAACTGTAAAGCATTAAGCAATGATTTATGTGA 1932
Db 1267 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1208
QY 1873 ATGAACTGTCATGTGATTTAAGCAATTTATCGCTAACAATGATTTATGTGA 1932
Db 1207 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1148
QY 1933 CAAGTCAGATTTCAAGTCATCAACAAGAGAGCTGAAGTGAAGTGAAGTAA 1992
Db 1147 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1088
QY 1993 AAAGAGCAGCATGAGAG 2010
Db 1087 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR

```

RESULT 2

```

US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHETTLINGER, F.
; TITLE OF INVENTION: FALKNER, F. G.
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)836-4109
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:

```

LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgtl-Fls
US-08-232-463-14

Query Match
Best Local Similarity 4.8%; Pred. No. 0.053;
Matches 15; Conservative 173; Mismatches 126; Indels 0; Gaps 0;

```

QY 739 GAGTATGCAATAGAGATAGACATTCAGAGTTTACTGACAGCTACAAAGTATTC 798
Db 1035 GAGCTTGCTGCGACGTCGAGGAGCTGCGATTTTTTTTTTTTTTTTTTTTTT 1094
QY 799 TTATAGTCCATTTCTTACTTCAATTTGATTTGTTGTTGTTGTTGTTGTTG 858
Db 1095 TTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1154
QY 859 ACTTGGGGTTTCATGAGAGCCGAGTATGATATCCATATTTCTGCTAATTTTCA 918
Db 1155 TTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1214
QY 919 CATCTCAACAGTTAAGATAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 978
Db 1215 TTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1274
QY 979 AGCTGTAATGTCGAAGTATGTTCAATTTTACTGTAATGTTGTTGTTGTTG 1038
Db 1275 TTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1334
QY 1039 TAATCCTTGCATC 1052
Db 1335 TTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1348

```

RESULT 3

```

US-08-973-462-2/c
; Sequence 2, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DROULHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCP
; CURRENT APPLICATION NUMBER: US/08/973,462B
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5361
; TYPE: DNA
; ORGANISM: P. falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(5361)
; US-08-973-462-2

```

Query Match
Best Local Similarity 1.9%; Score 40.2; DB 4; Length 5361;
Matches 102; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

```

QY 792 TGATGCTTTAGTCCATTTCTTACTTCAATTTGATTTGTTGTTGTTGTTG 851
Db 3984 TAATATATCTTCTTCAAGATTTTAAATGATTTTCTTCTTCTTCTTCTT 3925
QY 852 GCTTTGACTTTGGGGTTTCATGAGCTGAGTATGATGATGATGATGATGAT 911

```

05-09-41-150-3/C

Sequence 3, Application US/09741150
Patent No. 6436689
GENERAL INFORMATION:
APPLICANT: GUEGLER, Karl et al
TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
FILE REFERENCE: CL000968
CURRENT APPLICATION NUMBER: US/09/741,150
CURRENT FILING DATE: 2000-12-21
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 112132
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(112132)
OTHER INFORMATION: n = A,T,C or G
US-09-741-150-3

Query Match
Best Local Similarity 49.3%; Score 38.6; DB 4; Length 112132;
Matches 101; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 621 ATTTTGTTCCTTAATCAGGAACGTTTCAGGAAAGTATATAGCATGACAGGACTAG 680
DB 66017 ATTTTGTTCCTTTTCTTAACCAATTTTCAGGTGCTCTTATGAAATATACGAAACT 65958
QY 681 ATGTGAAGACCGAAGCTTCAATGCTTCTTTAAATCTTCAAGATTAGACTCTGAGGA 740
DB 65957 AGTACTACCCAGGCTCTTCTTCTTAAATGACTTCCTCAATATAGTCTCTCAAAA 65898
QY 741 GATGCAATFAGGATGAGACATGACAGTTTAATTGACGATCAAACTGATTCCT 800
DB 65897 ATTAAGTAATGGAATGACAGTGTATATGAGAACTTATATAAGACTGATTTAATAATA 65838
QY 801 TAGTCCCATTTCTCTTACTTTCATTT 825
DB 65837 AATCCTTCTTGTGCTTAGATTAATTT 65813

RESULT 7
US-08-991-677-11/C
Sequence 11, Application US/08991677A
Patent No. 6252135
GENERAL INFORMATION:
APPLICANT: Chiang, Vincent L
APPLICANT: Carriaway, Daniel T
APPLICANT: Smeltzer, Richard H
TITLE OF INVENTION: Production of Strylingyl Lignin in Gymnosperms
FILE REFERENCE: 50617
CURRENT APPLICATION NUMBER: US/08/991,677A
CURRENT FILING DATE: 1997-12-16
EARLIER APPLICATION NUMBER: 60/033,381
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 11
LENGTH: 2251
TYPE: DNA
ORGANISM: Pinus taeda
US-08-991-677-11

Query Match
Best Local Similarity 1.8%; Score 38.2; DB 4; Length 2251;
Matches 97; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 1023 TCTTCTTTAGGAGCTTAATCCCTTTCATCAGTTCAGCAGAGGATTACCAAGACA 1082
DB 560 TATATTTTAAACATCCATACATACATACATACATATATTTACTTAAAAAATCCCAAAATCA 501

QY 1083 AGTATTATACAGATATATCCATATATACACAAATATGCTGAGTGACCAAGAAATAC 1142
DB 500 CTATTTAAACACCCCTTATCCATATAAATAAGATACGTTATATGATTCGACATATAA 441
QY 1143 ACCAAATTCATCAGTTCAGTAACTTAATGCTGAGGTAAAGTATGATTCGAGAT 1202
DB 440 AACTAATGATTAATGAGACGACAAACATCTTCTTAAAGATAGACTTCAAGTTTGT 381
QY 1203 TATTTGACAGCTT 1217
DB 380 AGTTTCAAAATTT 366

RESULT 8
US-08-325-547-9
Sequence 9, Application US/08325547
Patent No. 5736378
GENERAL INFORMATION:
APPLICANT: Elder, John R.
APPLICANT: Talbot, Randy L.
TITLE OF INVENTION: NUCLEOTIDE SEQUENCE AND OPEN READING
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Spensley Horn Judas & Lubitz
STREET: 1880 Century Park East, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/325,547
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/759,570
FILING DATE: 12-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: Wetherell, Jr., Ph.D., John R.
REGISTRATION NUMBER: 31,678
REFERENCE/DOCKET NUMBER: PD-2000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 9472 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: 34
FEATURE:
NAME/KEY: CDS
LOCATION: join(627..1976, 1868..5239, 5235..5987, 5991
LOCATION: ..6224, 6264..8831, 6710..6913)
US-08-325-547-9

Query Match
Best Local Similarity 1.8%; Score 38; DB 1; Length 9472;
Matches 59; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 1070 GATTACCAAGACAGATATTATTAACAGATATGCTCATATATACAAATATGCTGAGGTG 1129
DB 8484 GATTACCAACAAAGTTTATGAAATATATGATGACATAGAACAAATATATATGACAGGG 8543
QY 1130 ACCACAAGATACACCAATTTACTCAATTGCAAG 1163

Db 8544 AAAACAGGATACACAACTTACAAAGTGGCAAG 8577

RESULT 9

US-09-134-001C-1438/C
Sequence 1438, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 1438
LENGTH: 522
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1438

Query Match 1.7%; Score 37.2; DB 4; Length 522;
Best Local Similarity 51.2%; Pred. No. 0.67; Mismatches 83; Indels 0; Gaps 0;
Matches 87; Conservative 0;

QY 694 ACTTCAATGCTTCTTTAAATCTCAAGATTAGACTCTTGAGAGTATGCAATAGA 753
DB 170 ACTCCATTTTAAATTTTATAGTATCAATAAATGAAATTTTATTCATTAAGCAATATT 111
QY 754 GATGAGACAAATGCAAGTTTACTTCAAGTACAAATGATTCCTTACTGCTTCC 813
DB 110 AAAATAGCAATGCTATACAAAGTAAAGTGTCCAAAAGCACCACATTAATCTCATTA 51
QY 814 TTACTTCATTTGATGATGTTGTGATCTTCGTTAGCTTGA 863
DB 50 TTAAATACCGTTTGATGATTAATTAAGACTTCAATGATACCTT 1

RESULT 10

US-08-811-682-18
Sequence 18, Application US/08811682
Patent No. 6331616
GENERAL INFORMATION:
APPLICANT: Tompkins, Wayne A.F.
APPLICANT: Tompkins, Mary B.
APPLICANT: Yang, Joo-Sung
TITLE OF INVENTION: Feline Immunodeficiency Virus Clone
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell Seltzer Park & Gibson
STREET: PO Drawer 34009
CITY: Charlotte
STATE: NC 6331616th Carolina
COUNTRY: USA
ZIP: 28234
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/811,682
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bennett, Virginia C.
REGISTRATION NUMBER: 37,092
REFERENCE/DOCKET NUMBER: 5051-332

TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-420-2200
TELEFAX: 919-881-3175
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2556 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-811-682-18

Query Match 1.7%; Score 36.4; DB 4; Length 2556;
Best Local Similarity 61.7%; Pred. No. 2.5; Mismatches 36; Indels 0; Gaps 0;
Matches 58; Conservative 0;

QY 1070 GATTACCAAGACATTTATACAGATTAATGTCATATATCAAAATATGCTGAGGTG 1129
DB 2209 GATTACCAAAAGTTTATGAAATATATATGACATAGCAAAATATATGTACAAAGG 2268
QY 1130 ACCACAGATACACCAATTAATCTCAATTCGCAAG 1163
DB 2269 AAAAAGGATACCAACTTACAAAGTGGCAAG 2302

RESULT 11

US-08-482-090-14
Sequence 14, Application US/08482090
Patent No. 5820869

GENERAL INFORMATION:
APPLICANT: Wasmoen, Terri
APPLICANT: Chu, Hsien-Jue
APPLICANT: Chavez, Lloyd
TITLE OF INVENTION: Recombinant Raccoon Fox Viruses and
TITLE OF INVENTION: Their Use as an Effective Vaccine Against Feline
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,090
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Schaffer, Robert
REGISTRATION NUMBER: 31,194
REFERENCE/DOCKET NUMBER: 0632/0B170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)527-7700
TELEFAX: (212)753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 3225 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Feline immunodeficiency virus
INDIVIDUAL ISOLATE: NCSU-1
POSITION IN GENOME:
MAP POSITION: 1-3225

UNITS: bp
US-08-482-090-14

Query Match 1.7%; Score 36.4; DB 1; Length 3225;
Best Local Similarity 61.7%; Pred. No. 2.8;
Matches 58; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1070 GATTACCAAGACAGTATTATACAGATATATGTCATATATCAAAATATGCTGAGGTG 1129
DB 2237 GATTACCAAGACAGTATTATGAAATATATATGACATAGAACCAATATATGTCAGAGG 2296
QY 1130 ACCACAGAGATACCAATATATACCAATATGCAAG 1163
DB 2297 AAACAGGATACCAATATACCAATGCGAG 2330

RESULT 12
US-08-481-700B-7
Sequence 7, Application US/08481700B
Patent No. 5849303

GENERAL INFORMATION:
APPLICANT: Wasmoen, Terri
APPLICANT: Chu, Hsien-Jue
TITLE OF INVENTION: Recombinant Feline Immunodeficiency
TITLE OF INVENTION: Vaccines Produced in Baculovirus and Their Use Against
TITLE OF INVENTION: Feline Immunodeficiency Virus Infection
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/481,700B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Schaffer, Robert
REGISTRATION NUMBER: 31,194
REFERENCE/DOCKET NUMBER: 0623/08168
TELEPHONE: (212)527-7700
TELEFAX: (212)753-6237
TELEX: 236687

INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 3225 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: feline immunodeficiency virus
INDIVIDUAL ISOLATE: NCSU-1
POSITION IN GENOME:
MAP POSITION: 1-3225
UNITS: bp
US-08-481-700B-7

Query Match 1.7%; Score 36.4; DB 2; Length 3225;
Best Local Similarity 61.7%; Pred. No. 2.8;
Matches 58; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1070 GATTACCAAGACAGTATTATACAGATATATGTCATATATCAAAATATGCTGAGGTG 1129
DB 2237 GATTACCAAGACAGTATTATGAAATATATATGACATAGAACCAATATATGTCAGAGG 2296

QY 1130 ACCACAGAGATACCAATATATACCAATATGCAAG 1163
DB 2297 AAACAGGATACCAATATACCAATGCGAG 2330

RESULT 13
US-09-007-383-14
Sequence 14, Application US/09007383A
Patent No. 5989562

GENERAL INFORMATION:
APPLICANT: Wasmoen, Terri
APPLICANT: Chu, Hsien-Jue
APPLICANT: Chavez, Lloyd
TITLE OF INVENTION: Recombinant Raccoon Pox Viruses And
TITLE OF INVENTION: Their Use As An Effective Vaccine Against Feline
TITLE OF INVENTION: Immunodeficiency Virus Infection
FILE REFERENCE: 0632/18170 US1
CURRENT APPLICATION NUMBER: US/09/007,383A
CURRENT FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 08/482,090
NUMBER OF SEQ ID NOS: 16
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 14
LENGTH: 3225
TYPE: DNA
ORGANISM: Feline Immunodeficiency Virus
US-09-007-383-14

Query Match 1.7%; Score 36.4; DB 2; Length 3225;
Best Local Similarity 61.7%; Pred. No. 2.8;
Matches 58; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1070 GATTACCAAGACAGTATTATACAGATATATGTCATATATCAAAATATGCTGAGGTG 1129
DB 2237 GATTACCAAGACAGTATTATGAAATATATATGACATAGAACCAATATATGTCAGAGG 2296
QY 1130 ACCACAGAGATACCAATATATACCAATATGCAAG 1163
DB 2297 AAACAGGATACCAATATACCAATGCGAG 2330

RESULT 14
US-08-811-682-1
Sequence 1, Application US/08811682
Patent No. 6331616

GENERAL INFORMATION:
APPLICANT: Tompkins, Wayne A.F.
APPLICANT: Tompkins, Mary B.
APPLICANT: Yang, Joo-Sung
TITLE OF INVENTION: Feline Immunodeficiency Virus Clone
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell Seltzer Park & Gibson
STREET: PO Drawer 34009
CITY: Charlotte
STATE: No. 6331616th Carolina
COUNTRY: USA
ZIP: 28234

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/811,682
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bennett, Virginia C.
REGISTRATION NUMBER: 37,092
REFERENCE/DOCKET NUMBER: 5051-332

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 18:35:35 ; Search time 23 Seconds
(without alignments)
420.876 Million cell updates/sec

Title: US-09-870-406A-33

Perfect score: 1724

Sequence: 1 MALSMERGFSGCFKAPNP.....STICKSYETRPWDLTGQFIG 329

Scoring table: BIOSUM62

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued_Patents_AA: *
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/PTUUS.COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110.5	6.4	514	US-09-182-859-6	Sequence 6, Appl1
2	110.5	6.4	514	US-09-264-097-8	Sequence 8, Appl1
3	110.5	6.4	514	US-09-672-459-6	Sequence 6, Appl1
4	110.5	6.4	515	US-09-291-023A-17	Sequence 17, Appl1
5	110.5	6.4	515	US-09-537-168-8	Sequence 8, Appl1
6	110.5	6.4	549	US-08-720-899-6	Sequence 6, Appl1
7	110.5	6.4	549	US-08-459-610-6	Sequence 6, Appl1
8	110.5	6.4	549	US-08-343-804-6	Sequence 6, Appl1
9	110.5	6.4	549	US-08-687-399-6	Sequence 6, Appl1
10	110.5	6.4	549	US-08-600-908A-6	Sequence 6, Appl1
11	110.5	6.4	549	US-08-683-838A-6	Sequence 6, Appl1
12	110.5	6.4	549	US-09-636-252A-6	Sequence 6, Appl1
13	110.5	6.4	548	US-08-468-700-37	Sequence 37, Appl1
14	110.5	6.4	548	US-08-645-971-5	Sequence 35, Appl1
15	110.5	6.4	548	US-08-468-220-35	Sequence 35, Appl1
16	110.5	6.4	548	US-08-468-698-35	Sequence 35, Appl1
17	110.5	6.4	548	US-08-704-706A-37	Sequence 37, Appl1
18	110.5	6.4	548	US-08-890-383-6	Sequence 6, Appl1
19	110.5	6.4	548	US-08-914-619A-6	Sequence 6, Appl1
20	110.5	6.4	548	US-08-985-659-38	Sequence 38, Appl1
21	110.5	6.4	548	US-08-194-664A-35	Sequence 35, Appl1
22	110.5	6.4	548	PCT-US94-01553A-35	Sequence 35, Appl1
23	110.5	6.4	548	PCT-US95-10426-35	Sequence 35, Appl1
24	110.5	6.4	727	US-09-134-001C-4067	Sequence 4067, Ap
25	89.5	5.2	855	US-08-619-554-6	Sequence 6, Appl1
26	89	5.2	369	US-08-181-629A-4	Sequence 4, Appl1
27	88.5	5.1	514	US-08-600-656-3	Sequence 3, Appl1

28	88.5	5.1	514	US-09-170-670-3	Sequence 3, Appl1
29	88.5	5.1	514	US-09-193-068-3	Sequence 3, Appl1
30	88.5	5.1	514	US-09-183-412-3	Sequence 3, Appl1
31	88.5	5.1	514	US-09-334-191A-3	Sequence 3, Appl1
32	88.5	5.1	514	US-09-290-734-3	Sequence 3, Appl1
33	88.5	5.1	573	US-09-134-001C-4813	Sequence 4813, Ap
34	87.5	5.0	1168	US-09-134-001C-4452	Sequence 4452, Ap
35	87	5.0	1168	US-08-448-1170-6	Sequence 6, Appl1
36	87	5.0	1168	US-08-961-803-8	Sequence 8, Appl1
37	85.5	5.0	386	US-08-968-751-6	Sequence 6, Appl1
38	85	4.9	748	US-08-408-318-2	Sequence 2, Appl1
39	85	4.9	748	US-08-839-164-2	Sequence 8, Appl1
40	85	4.9	748	US-09-087-465-8	Sequence 237, App
41	84.5	4.9	636	US-09-554-805-237	Sequence 17, Appl1
42	84	4.9	425	US-08-615-170-17	Sequence 3, Appl1
43	84	4.9	464	US-08-970-134-3	Sequence 4511, Ap
44	82	4.8	502	US-09-134-001C-4511	Sequence 48, Appl1
45	82	4.8	906	US-08-630-916A-48	

ALIGNMENTS

```
RESULT 1
US-09-182-859-6
; Sequence 6, Application US/09182859
; Patent No. 6143708
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; APPLICANT: Borchert, Torben
; APPLICANT: Bisgaard-Frantzen, Henrik
; TITLE OF INVENTION: Alpha-Amylase Mutants
; FILE REFERENCE: 4796.204-US
; CURRENT FILING DATE: US/09/182, 859
; EARLIER FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 0515/96
; EARLIER FILING DATE: 1996-04-30
; EARLIER APPLICATION NUMBER: 0712/96
; EARLIER FILING DATE: 1996-06-28
; EARLIER APPLICATION NUMBER: 0775/96
; EARLIER FILING DATE: 1996-07-11
; EARLIER APPLICATION NUMBER: 1263/96
; EARLIER FILING DATE: 1996-11-08
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
US-09-182-859-6
Query Match 6.4%; Score 110.5; DB 4; Length 514;
Best Local Similarity 19.4%; Pred. No. 0.0027;
Matches 68; Conservative 50; Mismatches 120; Indels 113; Gaps 16;
QY 50 SYKRAEALRETRRIYLEPSHLOEKSSMTGCKGKLELMARKSSKIRLRSMATEN 109
DB 195 NYDVLMTADLDMDHDEVTLEKNMGKMYNTNIDG-----FRIDAVKH 238
QY 110 ETMOVF-DFAQFMEPEYDPIFCANFSTWVNIYVLDJNPLHOTQDT-----Y 159
DB 239 IKFSFEPWMLVYRSQGTCKPLFTYGEYMS-----YDINKLHNITTDGMSLFDAPL 291
QY 160 QDKYI-----NKIMSYHKAEYF-----PMGGLKLG 186
DB 292 HNKRYTASGSGARDMRTLMTNTLTKDQPLTAVFVDNDHDEPGALQSWDPW----- 345
QY 187 ESIFESPFLVWMTSRSSSEKHKALFSAFLVEYQWLEMTIOVREMPSHVRANCEA-- 244
DB 346 -----FKPLA-YAFLITQEGYPCVF--YGDYIGIPQYNISLSKIDPLIARDYAYG 397
QY 245 -QHRYLT-----WRDQ---KDPGHGLKRL-----VGEAKAKELLRDLFNGV 283
```

Db 398 TQHDYLDHSDIIGWTRREGTEKPGSLALITDGPCKMIVYQKHAGKVFYDLTGNS 457
 QY 284 DELGKTFTIDYEPYQTEDEGTVS-----DKRSIIIGSKSYEMPMWLTGOFI 328
 Db 458 DVTIINS--DGMGEFKVNGSGSVWVPRKTYSTIARPIITTRPW--TGEFV 504

RESULT 2
 US-09-264-097-8
 ; Sequence 8, Application US/09264097
 ; Patent No. 6287826
 ; GENERAL INFORMATION:
 ; APPLICANT: No. 6287826man, Barrie Edmund
 ; APPLICANT: Hendriksen, Hanne Vang
 ; TITLE OF INVENTION: Enzymatic Preparation of Glucose Syrup
 ; FILE REFERENCE: 5278.200-US
 ; CURRENT APPLICATION NUMBER: US/09/264.097
 ; EARLIER FILING DATE: 1999-03-08
 ; EARLIER APPLICATION NUMBER: PA 0321/98
 ; EARLIER FILING DATE: 1998-03-09
 ; EARLIER APPLICATION NUMBER: 60/079.209
 ; EARLIER FILING DATE: 1998-03-24
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 8
 ; LENGTH: 514
 ; TYPE: PRT
 ; ORGANISM: Bacillus
 ; US-09-264-097-8

Query Match 6.4%; Score 110.5; DB 4; Length 514;
 Best Local Similarity 19.4%; Pred. No. 0.0027;
 Matches 68; Conservative 50; Mismatches 120; Indels 113; Gaps 16;
 QY 50 SYKEFAESALEETRRKRIYLEPSHLOEKYSMTGIDGKTELQMLAFKSKIRLSMAIEN 109
 Db 195 NYDIYAMADDMHPEVYTELKNMGKIVYNTNIDG-----FRIDAVKH 238
 QY 110 ETMOVF-DEAGEMEPEYDPIFCANFTSTNVAIVLDLNPRLHQLTDQD-----Y 159
 Db 239 IKSEFPDWLSTYRSQGRKPLFTVGEYWS-----YDINKLNIYTKTGTSMLDAPL 291
 QY 160 QDKRY-----NKIMSIYHKAETP-----PMGKILNG 186
 Db 292 HNKFTYASKSGAFDMRTLTMTLMDQPLAVTFDNDHTEPGQALQSWDPW----- 345
 QY 187 ESIKFSPLYMTREFSSSKKHKALSAFLEYQANLEMTIOVREEMPSHVANCEA-- 244
 Db 346 -----FKPLA-YAFILTRQGYPCVF--YGDYIGIPOYNIPSLKSIDPLIARRDYAIG 397
 QY 245 -QHKYLT-----WRAO---KDPGHGLKRL-----VGEAKAKELRDPLFNGV 283
 Db 398 TQHDYLDHSDIIGWTRREGTEKPGSLALITDGPCKMIVYQKHAGKVFYDLTGNS 457
 QY 284 DELGKTFTIDYEPYQTEDEGTVS-----DKRSIIIGSKSYEMPMWLTGOFI 328
 Db 458 DVTIINS--DGMGEFKVNGSGSVWVPRKTYSTIARPIITTRPW--TGEFV 504

RESULT 3
 US-09-672-459-6
 ; Sequence 6, Application US/09672459
 ; Patent No. 6436888
 ; GENERAL INFORMATION:
 ; APPLICANT: Svendsen, Allan
 ; APPLICANT: Borchert, Torben
 ; APPLICANT: Bisgard-Frantzen, Henrik
 ; TITLE OF INVENTION: Alpha-Amylase Mutants
 ; FILE REFERENCE: 4796.204-US
 ; CURRENT APPLICATION NUMBER: US/09/672.459
 ; CURRENT FILING DATE: 2000-09-28
 ; PRIOR APPLICATION NUMBER: 09/182.859

;; PRIOR FILING DATE: 1998-10-29
 ;; PRIOR APPLICATION NUMBER: 0515/96
 ;; PRIOR FILING DATE: 1996-04-30
 ;; PRIOR APPLICATION NUMBER: 0712/96
 ;; PRIOR FILING DATE: 1996-06-28
 ;; PRIOR APPLICATION NUMBER: 0775/96
 ;; PRIOR FILING DATE: 1996-07-11
 ;; PRIOR APPLICATION NUMBER: 1263/96
 ;; PRIOR FILING DATE: 1996-11-08
 ;; NUMBER OF SEQ ID NOS: 37
 ;; SOFTWARE: FastSeq for Windows Version 3.0
 ;; SEQ ID NO 6
 ;; LENGTH: 514
 ;; TYPE: PRT
 ;; ORGANISM: Bacillus stearothermophilus
 ; US-09-672-459-6

Query Match 6.4%; Score 110.5; DB 4; Length 514;
 Best Local Similarity 19.4%; Pred. No. 0.0027;
 Matches 68; Conservative 50; Mismatches 120; Indels 113; Gaps 16;

QY 50 SYKEFAESALEETRRKRIYLEPSHLOEKYSMTGIDGKTELQMLAFKSKIRLSMAIEN 109
 Db 195 NYDIYAMADDMHPEVYTELKNMGKIVYNTNIDG-----FRIDAVKH 238
 QY 110 ETMOVF-DEAGEMEPEYDPIFCANFTSTNVAIVLDLNPRLHQLTDQD-----Y 159
 Db 239 IKSEFPDWLSTYRSQGRKPLFTVGEYWS-----YDINKLNIYTKTGTSMLDAPL 291
 QY 160 QDKRY-----NKIMSIYHKAETP-----PMGKILNG 186
 Db 292 HNKFTYASKSGAFDMRTLTMTLMDQPLAVTFDNDHTEPGQALQSWDPW----- 345
 QY 187 ESIKFSPLYMTREFSSSKKHKALSAFLEYQANLEMTIOVREEMPSHVANCEA-- 244
 Db 346 -----FKPLA-YAFILTRQGYPCVF--YGDYIGIPOYNIPSLKSIDPLIARRDYAIG 397
 QY 245 -QHKYLT-----WRAO---KDPGHGLKRL-----VGEAKAKELRDPLFNGV 283
 Db 398 TQHDYLDHSDIIGWTRREGTEKPGSLALITDGPCKMIVYQKHAGKVFYDLTGNS 457
 QY 284 DELGKTFTIDYEPYQTEDEGTVS-----DKRSIIIGSKSYEMPMWLTGOFI 328
 Db 458 DVTIINS--DGMGEFKVNGSGSVWVPRKTYSTIARPIITTRPW--TGEFV 504

RESULT 4
 US-09-291-023A-17
 ; Sequence 17, Application US/09291023A
 ; Patent No. 6309871
 ; GENERAL INFORMATION:
 ; APPLICANT: Oultup, Helle
 ; APPLICANT: Borchert, Torben
 ; APPLICANT: Nielsen, Bjarne
 ; APPLICANT: Nielsen, Vibeke
 ; APPLICANT: Hoeck, Lisbeth
 ; TITLE OF INVENTION: Polypeptides Having Alkaline Alpha-amylase Activity And Nuclei
 ; FILE REFERENCE: 5821.010-US
 ; CURRENT APPLICATION NUMBER: US/09/291.023A
 ; CURRENT FILING DATE: 1999-04-13
 ; PRIOR APPLICATION NUMBER: DK 1999 00438
 ; PRIOR FILING DATE: 1999-03-31
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 17
 ; LENGTH: 515
 ; TYPE: PRT
 ; ORGANISM: Bacillus
 ; US-09-291-023A-17

Query Match 6.4%; Score 110.5; DB 4; Length 515;
 Best Local Similarity 19.4%; Pred. No. 0.0027;

	Matches	68; Conservative	50;	Mismatches	120; Indels	113; Gaps	16;
QY	50	SYKEPAESALEETRRKRIYLERSHLOEKYSMTGLDGKTQLOMLAFKSKIRLLRSMALLEN					109
Db	195	NYDLIADMLADLMDDPREVVELTKLNNGKWIVNTTINIDG-----FRIDAVKH					238
QY	110	ETMQVF-DEAGFMERPEVDPIFCANEFISTVINVIYLDNLPHLTQDQT-----Y					159
Db	239	IKFSFFPDWLISYVSQSOTGRPLEFTVEGYMS-----YDKLNHNITFTGTMSLPDAPL					291
QY	160	ODKY-----NKIMSIHYKAEP-----PMGGKLIG					186
Db	292	HNKITYTSKSGGARDMRTLNTLTKKDOPTLAIVTVDNHDTEPGALOSWDPM-----					345
QY	187	ESIAFSPLYVMTRFSSSEKKHKALLFSALEYQAEMLTIOVREMEPSHRANCEA--					244
Db	346	----FKRLA-YARILTRQEGYPCVF--YGDIYGIPOYNISLKSXIDPILARDYAYG					397
QY	245	-QHXYLT-----WRAQ---KDPGHGLIKRL-----VGEAKAKELLDFLENGV					283
Db	398	TQHOYLDHSIDIIGMTRREGTEKPGSGIALALLITDGPGSKMVMVVGQKHAGKVYDLTGNRS					457
QY	284	DELCTKNFLIDFPFYQTEDGVVS-----DKRSITGKSYERRPMULTQOFI					328
Db	458	DTYVINS--DSMGEEKVNGSVSVAWPARKTTVTSTARITTPRM--TSEFV					504

```

RESULT 5
US-09-537-168-8
; Sequence 8, Application US/09537168
; Patent No. 6410295
; GENERAL INFORMATION:
; APPLICANT: Andersen, Carsten
; APPLICANT: Bisgaard-Frantzen, Henrik
; APPLICANT: Jorgensen, Christel T.
; APPLICANT: Kjaerulff, Soren
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Alpha-Amylase Variants
; FILE REFERENCE: 5886.200-US
; CURRENT APPLICATION NUMBER: US/09/537,168
; CURRENT FILING DATE: 2000-03-29
; EARLIER APPLICATION NUMBER: PA 1999 00437
; EARLIER FILING DATE: 1999-03-30
; EARLIER APPLICATION NUMBER: 60/127,427
; EARLIER FILING DATE: 1999-04-01
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 515
; TYPE: PRF
; ORGANISM: Bacillus stearothermophilus
US-09-537-168-8

```

[illegible]

```
OY      245 --OKRYLT-----WRNO---KDPGGJLKRU-----VGEAAKELLRPFUNGV 283
        || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      398 TQMDYLDHSDIICWTEGGETERPGSLALLIDGCGSGMWAVYGQHAGKAFYDILTGNSR 457
        || | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      284 DELGCTRTFDYPEYOTEDGYVS-----DKRSITGKSTERRPMDLSQFT 328
        || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      458 DPTVLINS--DGWEGFVNGSGSVSWVPKRTYSTLARIPTRPW--TGEFV 504
        || | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 6
 US-08-720-899-6
 Sequence 6, Application US/08720899
 Patent No. 5753460
 GENERAL INFORMATION:
 APPLICANT: Bisgaard-Frantzen, Henrik
 APPLICANT: Borchert, Torben Vedel
 APPLICANT: Svendsen, Allan
 APPLICANT: Thellersen, Marianne
 APPLICANT: Van der Zee, Pia
 TITLE OF INVENTION: AMYLASE VARIANTS
 NUMBER OF SEQUENCES: 38
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 57534600 No. 5753460disk of No. 5753460th America, Inc.
 STREET: 405 Lexington Avenue, 64th Floor
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10174-6401
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/720,899
 FILING DATE: 10-OCT-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/343,804
 FILING DATE: 22-NOV-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Lowmney Dr., Karen A.
 REGISTRATION NUMBER: 31,274
 REFERENCE/DOCKET NUMBER: 4054.214-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-867-0123
 TELEFAX: 212-878-9655
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 549 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-720-899-6

```

Query Match      6.4%; Score 110.5; DB 1; Length 549;
Best Local Similarity 19.4%; Pred. No. 0.003;
Matches      66; Conservative      50; Mismatches 120; Indels 113; Gaps 16

QY      50  SYKEFAESAEETFRKRIIVLPERSHLDKYSMTGLDGTLELMQAFKSKIRLMSAILEN 109
Db       229 NYLYLMAYADMDHPVEVTELRKNMGWYVNTTNIDG-----FRLDVAKH 272
          | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      110 ETQGVF-DFAGFMEPEYDTPICAFNEFTSTNVNIVLVDLP,HO,LTQTD-----Y 158
          | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       273 IKRSEFPDMLSYVSQTKGRPLFTVGEYMS-----YDINKLHNYITKTDGMSLFDAPL 325
          | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      160 ODXYV-----NKINSIYKVAEFT-----PAGGLTLG 166
          | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       326 HNFETYASKSGCAFMDKRTLTMTLTLMKDDPTLVATVVDNHDHLEPGCALQSWDPH----- 379
          | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      187 ESIKFSPPLVMTTFRESSSEKKHKALESFALELYQAMLEMTIQVREMEPSHVNRANCEA-- 244
          | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Db 380 -----FKPLA-YAFILTRQEGYPCVF--YGDYGIPOYNIPSLKSKIDPLILARBYANG 431
QY 245 -OHKYL-----WRAO---KDPGHGLKRL-----VGEAKAKELLDLFLNGV 283
Db 432 TQHDYLDHSDILIGTRREGTEKFGSGGLALITDGGGSKMVMYVQKHAGKVFYDLTGNGRS 491
QY 284 DELGKTFTIDYFPEYQTEGTVS-----DKRSIIIGKSYETRPMDLTGOFI 328
Db 492 DTVTINS--DGWGEFKVNGSGSVWVPRKTVSTIARPIITTRPW--TGEFV 538

RESULT 7

US-08-459-610-6
; Sequence 6, Application US/08459610
; Patent No. 5801043
; GENERAL INFORMATION:
; APPLICANT: Bisgaard-Frantzen, Henrik
; APPLICANT: Borchert, Torben Vedel
; APPLICANT: Svendsen, Allan
; APPLICANT: Thellersen, Marianne
; APPLICANT: Van der Zee, Pia
; TITLE OF INVENTION: AMYLASE VARIANTS
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 58010430 No. 5801043disk of No. 5801043th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459, 610
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/343, 804
; FILING DATE: 22-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowmey Dr., Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4054, 214-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 549 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-459-610-6

Query Match 6.4%; Score 110.5; DB 1; Length 549;
Best Local Similarity 19.4%; Pred. No. 0.003;
Matches 68; Conservative 50; Mismatches 120; Indels 113; Gaps 16;

QY 50 SYKEFESALIEETRRKRIYLEPSHLOEKYSMTGLDGKTELOMLAFSSKIRLLRSMAIEN 109
Db 229 NYDIYLMYADLDMDHREYVTELKNGKMYVNTNIDG-----FLDAYKH 272
QY 110 ETMOVF-DEAGFMEPEYDPIFCANFTSTNVIIVLDLNLHQLTDQTD-----Y 159
Db 273 IKFSFPEWLSYVSQTKPLFTVGEYMS-----YDINKLHNYITKTDGMSLFDAPL 325
QY 160 QDKYY-----NKINSIYHKYAEFT-----PMGKLTG 186
Db 326 HNKFTYASKSGAFDMFTLMTNTLMKQDPLAVTFVNDHDEPQALQSWDPW----- 379

QY 187 ESIFESPFLVWMTFSSSKKHKALFSALTEYYOAMLEMTIOVREEMPSHVANCEA-- 244
Db 380 -----FKPLA-YAFILTRQEGYPCVF--YGDYGIPOYNIPSLKSKIDPLILARBYANG 431
QY 245 -OHKYL-----WRAO---KDPGHGLKRL-----VGEAKAKELLDLFLNGV 283
Db 432 TQHDYLDHSDILIGTRREGTEKFGSGGLALITDGGGSKMVMYVQKHAGKVFYDLTGNGRS 491
QY 284 DELGKTFTIDYFPEYQTEGTVS-----DKRSIIIGKSYETRPMDLTGOFI 328
Db 492 DTVTINS--DGWGEFKVNGSGSVWVPRKTVSTIARPIITTRPW--TGEFV 538

RESULT 8

US-08-343-804-6
; Sequence 6, Application US/08343804
; Patent No. 5830837
; GENERAL INFORMATION:
; APPLICANT: Bisgaard-Frantzen, Henrik
; APPLICANT: Borchert, Torben Vedel
; APPLICANT: Svendsen, Allan
; APPLICANT: Thellersen, Marianne
; APPLICANT: Van der Zee, Pia
; TITLE OF INVENTION: AMYLASE VARIANTS
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 58308370 No. 5830837disk of No. 5830837th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/343, 804
; FILING DATE: 22-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowmey Dr., Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4054, 214-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 549 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-343-804-6

Query Match 6.4%; Score 110.5; DB 2; Length 549;
Best Local Similarity 19.4%; Pred. No. 0.003;
Matches 68; Conservative 50; Mismatches 120; Indels 113; Gaps 16;

QY 50 SYKEFESALIEETRRKRIYLEPSHLOEKYSMTGLDGKTELOMLAFSSKIRLLRSMAIEN 109
Db 229 NYDIYLMYADLDMDHREYVTELKNGKMYVNTNIDG-----FLDAYKH 272
QY 110 ETMOVF-DEAGFMEPEYDPIFCANFTSTNVIIVLDLNLHQLTDQTD-----Y 159
Db 273 IKFSFPEWLSYVSQTKPLFTVGEYMS-----YDINKLHNYITKTDGMSLFDAPL 325
QY 160 QDKYY-----NKINSIYHKYAEFT-----PMGKLTG 186
Db 326 HNKFTYASKSGAFDMFTLMTNTLMKQDPLAVTFVNDHDEPQALQSWDPW----- 379
QY 187 ESIFESPFLVWMTFSSSKKHKALFSALTEYYOAMLEMTIOVREEMPSHVANCEA-- 244

Db 380 -----FKPLA-YAFLTRROEGYPCVF--YGDYGIPOYNIPSLKSKIDPLLIARDYAYG 431
QY 245 -QHKYLT-----WRAO---KDPGHGLKRL-----VEAKAKELLRDFLNGV 283
Db 432 TQHDYLDHSDIIGMTREGTEKPGSGLAALITDPPGSGKMYVGKOHGKVFYDLTGNS 491
QY 284 DELGKTFFIDYFPEYOTEDGTVS-----DKRSIIIGSKSYERPPMDLTGQFI 328
Db 492 DVTIINS--DGMGEFKVNGGSVSWVPKRTTVSTIARITTRPM--TGEFV 538

RESULT 9

US-08-687-399-6
Sequence 6, Application US/08687399
Patent No. 5928381
GENERAL INFORMATION:
APPLICANT: Toft, Annette H.
APPLICANT: Marcher, Dorte
APPLICANT: Pedersen, Hanne H.
APPLICANT: Nilsson, Thomas E.
TITLE OF INVENTION: A Combined Desizing and Bleaching
TITLE OF INVENTION: Process
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 59283810 No. 5928381disk of No. 5928381th America, Inc.
STREET: 405 Lexington Avenue, 64th floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,399
FILING DATE:
CLASSIFICATION: 008
ATTORNEY/AGENT INFORMATION:
NAME: Lambdlis, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4127,204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 549 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-687-399-6

Query Match 6.4%; Score 110.5; DB 2; Length 549;

Best Local Similarity 19.4%; Pred. No. 0.003; Indels 113; Gaps 16;

Matches 68; Conservative 50; Mismatches 120;

QY 50 SYKEFASALEETRRKRIYLEPSHLOEKYSMTGLDGTETLOMLAFKSKIRILRSMALIN 109
Db 229 NYDIAMVADLDMDPREVYTELKNGKMYVNTNIDG-----FRLDAVKH 272
QY 110 ETMOVF--DEAGFMEPEVDPIFCANFTSTNVNIVLDLNPLOLTDOTD-----Y 159
Db 273 IKSFEPDMLSYVNSQGKPLFTYGEYWS-----YDINKLHNYITKTDGMSLFDAPL 325
QY 160 QDKYY-----NKIMSIYHKYAEF-----PMGKILTG 186
Db 326 HNKFTYASKSGAFDMRTIMTLMKDQPLATVFNHNDTEPGQALQSWDPM----- 379
QY 187 ESIFESPVLWMTFRSSSKERKHALFSAFLEYOAMLEMTIQVEEMEPSHVANCA-- 244

Db 380 -----FKPLA-YAFLTRROEGYPCVF--YGDYGIPOYNIPSLKSKIDPLLIARDYAYG 431
QY 245 -QHKYLT-----WRAO---KDPGHGLKRL-----VEAKAKELLRDFLNGV 283
Db 432 TQHDYLDHSDIIGMTREGTEKPGSGLAALITDPPGSGKMYVGKOHGKVFYDLTGNS 491
QY 284 DELGKTFFIDYFPEYOTEDGTVS-----DKRSIIIGSKSYERPPMDLTGQFI 328
Db 492 DVTIINS--DGMGEFKVNGGSVSWVPKRTTVSTIARITTRPM--TGEFV 538

RESULT 10

US-08-600-908A-6
Sequence 6, Application US/08600908A
Patent No. 5989169
GENERAL INFORMATION:
APPLICANT: Svendsen, Allan
APPLICANT: Bisg rd-Frantzen, Henrik
APPLICANT: Borchert, Torben Vedel
TITLE OF INVENTION: Amylase Mutants
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 59891690 No. 5989169disk of No. 5989169th America, Inc.
STREET: 405 Lexington Avenue, 64th floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/600,908A
FILING DATE: 13-FEB-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Green, Reza
REGISTRATION NUMBER: 38,475
REFERENCE/DOCKET NUMBER: 4394,204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 549 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-600-908A-6

Query Match 6.4%; Score 110.5; DB 2; Length 549;

Best Local Similarity 19.4%; Pred. No. 0.003; Indels 113; Gaps 16;

Matches 68; Conservative 50; Mismatches 120;

QY 50 SYKEFASALEETRRKRIYLEPSHLOEKYSMTGLDGTETLOMLAFKSKIRILRSMALIN 109
Db 229 NYDIAMVADLDMDPREVYTELKNGKMYVNTNIDG-----FRLDAVKH 272
QY 110 ETMOVF--DEAGFMEPEVDPIFCANFTSTNVNIVLDLNPLOLTDOTD-----Y 159
Db 273 IKSFEPDMLSYVNSQGKPLFTYGEYWS-----YDINKLHNYITKTDGMSLFDAPL 325
QY 160 QDKYY-----NKIMSIYHKYAEF-----PMGKILTG 186
Db 326 HNKFTYASKSGAFDMRTIMTLMKDQPLATVFNHNDTEPGQALQSWDPM----- 379
QY 187 ESIFESPVLWMTFRSSSKERKHALFSAFLEYOAMLEMTIQVEEMEPSHVANCA-- 244
Db 380 -----FKPLA-YAFLTRROEGYPCVF--YGDYGIPOYNIPSLKSKIDPLLIARDYAYG 431
QY 245 -QHKYLT-----WRAO---KDPGHGLKRL-----VEAKAKELLRDFLNGV 283

Db 432 TQHDYLDHSDIIGTREGGTEKPSGSLAALITDGPSSKMMYVGKQAHGKVFYDLTGNS 491
QY 284 DELGTCTFDYFPEYQEDGTVS-----DKRSIIKSYETRPWDLTGQFI 328
Db 492 DTVTINS--DGWGEFKVNGGSVSWVPRKTTVTIARPIITRPW--TGEFV 538

RESULT 11

US-08-683-838A-6
Sequence 6, Application US/08683838A
Patent No. 6022724
GENERAL INFORMATION:
APPLICANT: Svendsen, Allan
APPLICANT: Bisg rd-Frantzen, Henrik
APPLICANT: Borchert, Torben Vedel
TITLE OF INVENTION: -Amylase Mutants
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6022724o No. 6022724disk of No. 6022724th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/683,838A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/600,908
FILING DATE: 13-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Green, Reza
REGISTRATION NUMBER: 38,475
REFERENCE/DOCKET NUMBER: 4394.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 549 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-683-838A-6

Query Match 6.4%; Score 110.5; DB 3; Length 549;
Best Local Similarity 19.4%; Pred. No. 0.003;
Matches 68; Conservative 50; Mismatches 120; Indels 113; Gaps 16;

QY 50 SYEPFASALEETFRKRIYLEPSHLQEKYSMTGLDGKTELQMLAFKSSKIRILRSMAIEN 109
Db 229 NYDIYMYADLDMDHPEVTELEKMGKMYVNTNIDG-----FRIDAVKH 272
QY 110 ETMQVF-DFAGFMEPEYDPIFCANFTSTNNVIVLDLNPILQILDQND-----Y 159
Db 273 IKSFPPDLSTYRSOTGRPLFTVGGEYS-----YDINKLNNYITKIDGTMSLDAPL 325
QY 160 QDKYY-----NKIMSIYHKAETP-----PMGSKLNG 186
Db 326 HNKFTYASKSGAFDMRTILMTLMKDQPTIATVFDNHDTEFGQALQSWDPM----- 379
QY 187 ESIKFSPILVMYTRFSSSEKHKALFSAPLEYQAWLEMTIOVREEMESHVANCEA-- 244
Db 380 -----FKPLA-YAFILTRQEGYPCVF--YGDYIGIPQYNIPLSKSIDILLARRDYAAG 431
QY 245 -QHKYLT-----WRAQ--KDPGHGLKRL-----VGEAKAKELLRLDFLNGV 283

Db 432 TQHDYLDHSDIIGTREGGTEKPSGSLAALITDGPSSKMMYVGKQAHGKVFYDLTGNS 491
QY 284 DELGTCTFDYFPEYQEDGTVS-----DKRSIIKSYETRPWDLTGQFI 328
Db 492 DTVTINS--DGWGEFKVNGGSVSWVPRKTTVTIARPIITRPW--TGEFV 538

RESULT 12

US-09-636-252A-6
Sequence 6, Application US/09636252A
Patent No. 6440716
GENERAL INFORMATION:
APPLICANT: Svendsen, Allan
APPLICANT: Bisgard-Frantzen, Henrik
APPLICANT: Borchert, Torben Vedel
TITLE OF INVENTION: Alpha-Amylase Mutants
FILE REFERENCE: 0776/1F216-052
CURRENT APPLICATION NUMBER: US/09/636,252A
CURRENT FILING DATE: 2000-08-10
PRIOR APPLICATION NUMBER: 08/683,838
NUMBER OF SEQ ID NOS: 16
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 6
LENGTH: 549
TYPE: PRT
ORGANISM: B. stearohermophilus
US-09-636-252A-6

Query Match 6.4%; Score 110.5; DB 4; Length 549;
Best Local Similarity 19.4%; Pred. No. 0.003;
Matches 68; Conservative 50; Mismatches 120; Indels 113; Gaps 16;

QY 50 SYEPFASALEETFRKRIYLEPSHLQEKYSMTGLDGKTELQMLAFKSSKIRILRSMAIEN 109
Db 229 NYDIYMYADLDMDHPEVTELEKMGKMYVNTNIDG-----FRIDAVKH 272
QY 110 ETMQVF-DFAGFMEPEYDPIFCANFTSTNNVIVLDLNPILQILDQND-----Y 159
Db 273 IKSFPPDLSTYRSOTGRPLFTVGGEYS-----YDINKLNNYITKIDGTMSLDAPL 325
QY 160 QDKYY-----NKIMSIYHKAETP-----PMGSKLNG 186
Db 326 HNKFTYASKSGAFDMRTILMTLMKDQPTIATVFDNHDTEFGQALQSWDPM----- 379
QY 187 ESIKFSPILVMYTRFSSSEKHKALFSAPLEYQAWLEMTIOVREEMESHVANCEA-- 244
Db 380 -----FKPLA-YAFILTRQEGYPCVF--YGDYIGIPQYNIPLSKSIDILLARRDYAAG 431
QY 245 -QHKYLT-----WRAQ--KDPGHGLKRL-----VGEAKAKELLRLDFLNGV 283
Db 432 TQHDYLDHSDIIGTREGGTEKPSGSLAALITDGPSSKMMYVGKQAHGKVFYDLTGNS 491
QY 284 DELGTCTFDYFPEYQEDGTVS-----DKRSIIKSYETRPWDLTGQFI 328
Db 492 DTVTINS--DGWGEFKVNGGSVSWVPRKTTVTIARPIITRPW--TGEFV 538

RESULT 13
US-08-468-700-37
Sequence 37, Application US/08468700
Patent No. 5736499
GENERAL INFORMATION:
APPLICANT: COLIN MITCHINSON
APPLICANT: CAROL A. REODADT
APPLICANT: TRACI H. ROEP
APPLICANT: LEIF P. SOLHEIM
TITLE OF INVENTION: MUTANT ALPHA-AMYLASE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International
STREET: 180 Kimball Way

CITY: South San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,700
FILING DATE: 06-JUN-1995
CLASSIFICATION: 252
ATTORNEY/AGENT INFORMATION:
NAME: Stone, Christopher
REGISTRATION NUMBER: 35,696
REFERENCE/DOCKET NUMBER: GC275
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 742-7555
TELEFAX: (415) 742-7217
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 548 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-468-700-37

Query Match 6.1%; Score 105; DB 1; Length 548;
Best Local Similarity 19.3%; Pred. No. 0.012;
Matches 68; Conservative 50; Mismatches 117; Indels 118; Gaps 17;
QY 50 SYKEFAESALEETRRRIYLEPSHLOEKSSMTGDKTELOMLAFKSKIRLLSMAIEN 109
DB 229 NYDYLMYADLDMDHPEVYTELKNMGKWVNTNIDG-----FRIDGLKH 272
QY 110 ETMOVF-DFAGMEPEYDPTIFCA-----NFTSTNNVIYVLDLNPRLQLDQ 156
DB 273 IKFSFPDWLSYVSQTKPLFTYGEWYSYDINKLHNTYTKTGMSLFLDA-PLH----- 326
QY 157 TDYODKYY-----NKIMSIYKVAETP-----PMGCKL 184
DB 327 ---NKFYTASKSGCAFDMRTLMTLTKKDQPLAVTFVDNDHDPARCSHGPRW----- 378
QY 185 TGESIKFSPPLVMTFRSSSKERKALFSALEYQAWLEMTIOVREEMPSHVANCEA 244
DB 379 -----FKPLA-YAFILTRQEGYPCVF--YGDYVGIPQVYNIPSLKSIDPILLARRDYA 428
QY 245 ---OHKYYLT-----WRAQ---KDPGHGLKRL-----VGEAKAKELLDFLNF 281
DB 429 YGTQHDYLDHSDIIGWTRREGVTEKRGSGLLALITDGAGRSKMWYVKQHAGKVFYDLTGN 488
QY 282 GVDLGTKTFIDYFPEYQTEDETVS-----DKRSIIKSYETRPMDLTGQFI 328
DB 489 RSDIYTTINS--DGWGEFVNGSGSVWVPRKTYVTIARPIITTRW--TGEFV 537

RESULT 14
US-08-645-971-5
Sequence 5, Application US/08645971
Patent No. 5763385
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Modified Alpha-Amylases Having Altered
NUMBER OF SEQUENCES: 5
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/645,971
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 548 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Bacillus stearothermophilus
US-08-645-971-5

Query Match 6.1%; Score 105; DB 1; Length 548;
Best Local Similarity 19.3%; Pred. No. 0.012;
Matches 68; Conservative 50; Mismatches 117; Indels 118; Gaps 17;
QY 50 SYKEFAESALEETRRRIYLEPSHLOEKSSMTGDKTELOMLAFKSKIRLLSMAIEN 109
DB 229 NYDYLMYADLDMDHPEVYTELKNMGKWVNTNIDG-----FRIDGLKH 272
QY 110 ETMOVF-DFAGMEPEYDPTIFCA-----NFTSTNNVIYVLDLNPRLQLDQ 156
DB 273 IKFSFPDWLSYVSQTKPLFTYGEWYSYDINKLHNTYTKTGMSLFLDA-PLH----- 326
QY 157 TDYODKYY-----NKIMSIYKVAETP-----PMGCKL 184
DB 327 ---NKFYTASKSGCAFDMRTLMTLTKKDQPLAVTFVDNDHDPARCSHGPRW----- 378
QY 185 TGESIKFSPPLVMTFRSSSKERKALFSALEYQAWLEMTIOVREEMPSHVANCEA 244
DB 379 -----FKPLA-YAFILTRQEGYPCVF--YGDYVGIPQVYNIPSLKSIDPILLARRDYA 428
QY 245 ---OHKYYLT-----WRAQ---KDPGHGLKRL-----VGEAKAKELLDFLNF 281
DB 429 YGTQHDYLDHSDIIGWTRREGVTEKRGSGLLALITDGAGRSKMWYVKQHAGKVFYDLTGN 488
QY 282 GVDLGTKTFIDYFPEYQTEDETVS-----DKRSIIKSYETRPMDLTGQFI 328
DB 489 RSDIYTTINS--DGWGEFVNGSGSVWVPRKTYVTIARPIITTRW--TGEFV 537

RESULT 15
US-08-468-220-35
Sequence 35, Application US/08468220
Patent No. 5824532
GENERAL INFORMATION:
APPLICANT: Antim, Richard L.
APPLICANT: Barnett, Christopher
APPLICANT: Mitchinson, Collin
APPLICANT: Power, Scott D.
APPLICANT: Regardt, Carol
APPLICANT: Solheim, Leif P.
TITLE OF INVENTION: Oxidatively Stable Alpha-Amylase
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESS: Genecore International, Inc.
STREET: 180 Kimball Way
CITY: South San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 MB
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,220
FILING DATE: 06-JUN-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/194,664
FILING DATE: 10-FEB-94

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 18:38:00 ; Search time 160 Seconds
(without alignments)

222.500 Million cell updates/sec

Title: US-09-870-406A-33

Perfect score: 1724

Sequence: 1 MALSMERFSGSCFKAPNP.....SIICKSYETRPMDLTGQFIG 329

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing:

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*

2: /cgn2_6/ptodata/2/pubpaa/PCR_NEW_PUB.pep:*

3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*

4: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*

5: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*

6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*

7: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*

8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*

10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*

11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*

12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*

13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*

14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1724	100.0	329	9	US-10-159-901-33
2	1724	100.0	329	9	US-10-159-901-34
3	1724	100.0	329	9	US-10-159-901-55
4	1724	100.0	329	9	US-09-870-406A-33
5	1724	100.0	329	9	US-09-870-406A-34
6	1724	100.0	329	9	US-09-870-406A-35
7	339.5	19.7	255	9	US-10-159-901-54
8	339.5	19.7	255	9	US-09-870-406A-54
9	215	12.5	257	9	US-10-159-901-38
10	215	12.5	257	9	US-10-159-901-52
11	215	12.5	257	9	US-09-870-406A-38
12	215	12.5	257	9	US-09-870-406A-52
13	210.5	12.2	257	9	US-10-159-901-36
14	210.5	12.2	257	9	US-10-159-901-50
15	210.5	12.2	257	9	US-09-870-406A-36
16	210.5	12.2	257	9	US-09-870-406A-50
17	209	12.1	262	9	US-10-159-901-51
18	209	12.1	262	9	US-09-870-406A-51
19	205	11.9	280	9	US-10-159-901-49

20	205	11.9	280	9	US-09-870-406A-49	Sequence 49, Appl
21	172.5	10.0	257	9	US-10-159-901-53	Sequence 53, Appl
22	172.5	10.0	257	9	US-09-870-406A-53	Sequence 53, Appl
23	158	9.2	236	9	US-10-159-901-35	Sequence 35, Appl
24	158	9.2	236	9	US-10-159-901-45	Sequence 45, Appl
25	158	9.2	236	9	US-09-870-406A-35	Sequence 35, Appl
26	158	9.2	236	9	US-09-870-406A-45	Sequence 45, Appl
27	145	8.4	235	9	US-10-159-901-46	Sequence 46, Appl
28	145	8.4	235	9	US-09-870-406A-46	Sequence 46, Appl
29	136	7.9	236	9	US-10-159-901-48	Sequence 48, Appl
30	136	7.9	236	9	US-09-870-406A-48	Sequence 48, Appl
31	122.5	7.1	241	9	US-10-159-901-37	Sequence 37, Appl
32	122.5	7.1	241	9	US-10-159-901-47	Sequence 47, Appl
33	122.5	7.1	241	9	US-09-870-406A-37	Sequence 37, Appl
34	122.5	7.1	241	9	US-09-870-406A-47	Sequence 47, Appl
35	121.5	7.0	490	9	US-10-159-901-40	Sequence 40, Appl
36	121.5	7.0	490	9	US-09-870-406A-40	Sequence 40, Appl
37	110.5	6.4	515	9	US-09-918-543-6	Sequence 6, Appl1
38	110.5	6.4	515	9	US-10-146-327-8	Sequence 8, Appl1
39	110.5	6.4	515	10	US-09-854-346-6	Sequence 6, Appl1
40	109.5	6.4	245	9	US-09-870-406A-41	Sequence 41, Appl
41	109.5	6.4	245	9	US-10-159-901-39	Sequence 39, Appl
42	105.5	6.1	248	9	US-10-159-901-42	Sequence 42, Appl
43	105.5	6.1	248	9	US-09-870-406A-39	Sequence 39, Appl
44	105.5	6.1	248	9	US-09-870-406A-42	Sequence 42, Appl
45	105.5	6.1	248	9	US-09-870-406A-42	Sequence 42, Appl

ALIGNMENTS

RESULT 1

US-10-159-901-33
Sequence 33, Application US/10159901
Publication No. US2003007335A1
GENERAL INFORMATION:
APPLICANT: LAGARIAS, JOHN
APPLICANT: KOICHI, TAKAYUKI
APPLICANT: FRANKENBERG, NICOLE
APPLICANT: GAMEYRA, GREGORY
APPLICANT: MONTGOMERY, BERONDA
TITLE OF INVENTION: LIGHT CONTROLLED GENE EXPRESSION UTILIZING HETEROLOGOUS PHYTOC
FILE REFERENCE: 407T-907731US
CURRENT APPLICATION NUMBER: US/10/159,901
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: 60/294,463
PRIOR FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 57
SOFTWARE: PatentIn version 3.0
SEQ ID NO 33
LENGTH: 329
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-10-159-901-33

Query Match 100.0%: Score 1724; DB 9; Length 329;
Best Local Similarity 100.0%: Pred. No. 3.4e-158;
Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALSMERFSGSCFKAPNPVYLISAPNKINFTLRKKRRLRVSAVSXFEASALE 60
|||||
DB 1 MALSMERFSGSCFKAPNPVYLISAPNKINFTLRKKRRLRVSAVSXFEASALE 60
|||||
QY 61 ETRKRIVLEPSHLQKYSYSGTGLDQTELOMAFSSKIRLRSAIENEMQVDFDGF 120
|||||
DB 61 ETRKRIVLEPSHLQKYSYSGTGLDQTELOMAFSSKIRLRSAIENEMQVDFDGF 120
|||||
QY 121 MEPEVDITFCANFTSTNNVIVLDPHLQTOGYODKYKIKSIYKVAETFPW 180
|||||
DB 121 MEPEVDITFCANFTSTNNVIVLDPHLQTOGYODKYKIKSIYKVAETFPW 180
|||||
QY 181 GGLTGESIKFFSPILVMTRESSSEKKKALFSAFLVYQAWLEMTJOVREMEPSHYRA 240
|||||

```

Db      181 GGLTGESIKFSPVMTFRSSSEKHKALFSAFLEYQAMLENTIOVREEMPSHVRA 240
QY      241 NCEAOKHYLTWRAQDPGGLIKRLVGAFAKAKELLROFLFNGVDELGTKEFDYFPEYOT 300
Db      241 NCEAOKHYLTWRAQDPGGLIKRLVGAFAKAKELLROFLFNGVDELGTKEFDYFPEYOT 300
QY      301 EDGTVSDKRSIIIGKSYETRPMDLTGQFIG 329
Db      301 EDGTVSDKRSIIIGKSYETRPMDLTGQFIG 329

```

RESULT 2

```

US-10-159-901-34
; Sequence 34, Application US/10159901
; Publication No. US20030073235A1
; GENERAL INFORMATION:
; APPLICANT: LAGARIAS, JOHN
; APPLICANT: KOICHI, TAKAYUKI
; APPLICANT: FRANKENBERG, NICOLE
; APPLICANT: GAMBETTA, GREGORY
; APPLICANT: MONTGOMERY, BERONDA
; TITLE OF INVENTION: LIGHT CONTROLLED GENE EXPRESSION UTILIZING HETEROLOGOUS PHYTOCHROME
; FILE REFERENCE: 407T-907731US
; CURRENT APPLICATION NUMBER: US/10/159,901
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: 60/294,463
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-159-901-34

```

Query Match Best Local Similarity 100.0%; Score 1724; DB 9; Length 329; Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      1 MALSMFSGISGCFKAPNPVLISAPNKINFTLRKKRRLRLVSAVSEFAESALE 60
Db      1 MALSMFSGISGCFKAPNPVLISAPNKINFTLRKKRRLRLVSAVSEFAESALE 60
QY      61 ETRKRIVLEPSHLOEKYSMTGLDGKTELQMLAFKSSKIRLRSMALENETMOVDFPAGF 120
Db      61 ETRKRIVLEPSHLOEKYSMTGLDGKTELQMLAFKSSKIRLRSMALENETMOVDFPAGF 120
QY      121 MEPEYDPIFCANFTSTNVNIVLDLNPLOLTDQTDYODKYNNKIMSTYHRYAETFPW 180
Db      121 MEPEYDPIFCANFTSTNVNIVLDLNPLOLTDQTDYODKYNNKIMSTYHRYAETFPW 180
QY      181 GGLTGESIKFSPVMTFRSSSEKHKALFSAFLEYQAMLENTIOVREEMPSHVRA 240
Db      181 GGLTGESIKFSPVMTFRSSSEKHKALFSAFLEYQAMLENTIOVREEMPSHVRA 240
QY      241 NCEAOKHYLTWRAQDPGGLIKRLVGAFAKAKELLROFLFNGVDELGTKEFDYFPEYOT 300
Db      241 NCEAOKHYLTWRAQDPGGLIKRLVGAFAKAKELLROFLFNGVDELGTKEFDYFPEYOT 300
QY      301 EDGTVSDKRSIIIGKSYETRPMDLTGQFIG 329
Db      301 EDGTVSDKRSIIIGKSYETRPMDLTGQFIG 329

```

RESULT 3

```

US-10-159-901-55
; Sequence 55, Application US/10159901
; Publication No. US20030073235A1
; GENERAL INFORMATION:
; APPLICANT: LAGARIAS, JOHN
; APPLICANT: KOICHI, TAKAYUKI
; APPLICANT: FRANKENBERG, NICOLE
; APPLICANT: GAMBETTA, GREGORY

```

```

; APPLICANT: MONTGOMERY, BERONDA
; TITLE OF INVENTION: LIGHT CONTROLLED GENE EXPRESSION UTILIZING HETEROLOGOUS PHYTOCHROME
; FILE REFERENCE: 407T-907731US
; CURRENT APPLICATION NUMBER: US/10/159,901
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: 60/294,463
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 55
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-159-901-55

```

Query Match Best Local Similarity 100.0%; Score 1724; DB 9; Length 329; Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      1 MALSMFSGISGCFKAPNPVLISAPNKINFTLRKKRRLRLVSAVSEFAESALE 60
Db      1 MALSMFSGISGCFKAPNPVLISAPNKINFTLRKKRRLRLVSAVSEFAESALE 60
QY      61 ETRKRIVLEPSHLOEKYSMTGLDGKTELQMLAFKSSKIRLRSMALENETMOVDFPAGF 120
Db      61 ETRKRIVLEPSHLOEKYSMTGLDGKTELQMLAFKSSKIRLRSMALENETMOVDFPAGF 120
QY      121 MEPEYDPIFCANFTSTNVNIVLDLNPLOLTDQTDYODKYNNKIMSTYHRYAETFPW 180
Db      121 MEPEYDPIFCANFTSTNVNIVLDLNPLOLTDQTDYODKYNNKIMSTYHRYAETFPW 180
QY      181 GGLTGESIKFSPVMTFRSSSEKHKALFSAFLEYQAMLENTIOVREEMPSHVRA 240
Db      181 GGLTGESIKFSPVMTFRSSSEKHKALFSAFLEYQAMLENTIOVREEMPSHVRA 240
QY      241 NCEAOKHYLTWRAQDPGGLIKRLVGAFAKAKELLROFLFNGVDELGTKEFDYFPEYOT 300
Db      241 NCEAOKHYLTWRAQDPGGLIKRLVGAFAKAKELLROFLFNGVDELGTKEFDYFPEYOT 300
QY      301 EDGTVSDKRSIIIGKSYETRPMDLTGQFIG 329
Db      301 EDGTVSDKRSIIIGKSYETRPMDLTGQFIG 329

```

RESULT 4

```

US-09-870-406a-33
; Sequence 33, Application US/09870406a
; Publication No. US20030104379A1
; GENERAL INFORMATION:
; APPLICANT: LAGARIAS, JOHN
; APPLICANT: KOICHI, TAKAYUKI
; APPLICANT: FRANKENBERG, NICOLE
; APPLICANT: GAMBETTA, GREGORY
; APPLICANT: MONTGOMERY, BERONDA
; TITLE OF INVENTION: HY2 FAMILY OF BILIN REDUCTASES
; FILE REFERENCE: 407T-907720US
; CURRENT APPLICATION NUMBER: US/09/870,406a
; PRIOR FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: 60/271,758
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 60/294,463
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 33
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-870-406a-33

```

Query Match Best Local Similarity 100.0%; Score 1724; DB 9; Length 329; Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MALSMFSGISGSCFAPNPVLLISAPNKINFTLRKKRRLVSAVSYPEFASALE 60
DB 1 MALSMFSGISGSCFAPNPVLLISAPNKINFTLRKKRRLVSAVSYPEFASALE 60
QY 61 ETRKRIVLEPSHLOEKYSMTGLDQKTELQMLAFKSSKIRLLRSMAIENETMQVDFAGF 120
DB 61 ETRKRIVLEPSHLOEKYSMTGLDQKTELQMLAFKSSKIRLLRSMAIENETMQVDFAGF 120
QY 121 MEPEYDPIFCANFTSTNNVIVLDLNPRLHQLDQDTYODKYKIMSIYHKYAEETPPW 180
DB 121 MEPEYDPIFCANFTSTNNVIVLDLNPRLHQLDQDTYODKYKIMSIYHKYAEETPPW 180
QY 181 GSKLTGESIKFSPVLMWTRFSSSEKHKALFSAFLEYQOAWLEMTIOVREMEPSHYRA 240
DB 181 GSKLTGESIKFSPVLMWTRFSSSEKHKALFSAFLEYQOAWLEMTIOVREMEPSHYRA 240
QY 241 NCEAOHKYLTWRAQKDPGHGLLKRVLGSAKAKELLRDLFNGVDELGKTFIDYPEPYOT 300
DB 241 NCEAOHKYLTWRAQKDPGHGLLKRVLGSAKAKELLRDLFNGVDELGKTFIDYPEPYOT 300
QY 301 EDGTVSDKRSITIGKSYETRPMDLTGQFTG 329
DB 301 EDGTVSDKRSITIGKSYETRPMDLTGQFTG 329

```

RESULT 5

```

US-09-870-406a-34
; Sequence 34, Application US/09870406A
; Publication No. US20030104379A1
; GENERAL INFORMATION:
; APPLICANT: LAGARIAS, JOHN
; APPLICANT: KOICHI, TAKAYUKI
; APPLICANT: FRANKENBERG, NICOLE
; APPLICANT: GAMBETTA, GREGORY
; APPLICANT: MONTGOMERY, BERONDA
; TITLE OF INVENTION: HY2 FAMILY OF BILLIN REDUCTASES
; FILE REFERENCE: 407T-907720US
; CURRENT APPLICATION NUMBER: US/09/870,406A
; PRIOR FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: 60/271,758
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 60/210,286
; PRIOR FILING DATE: 2000-06-08
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Arapidopsis thaliana
US-09-870-406a-34

```

Query Match

```

Best Local Similarity 100.0%; Score 1724; DB 9; Length 329;
Pred. No. 3,4e-158;
Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MALSMFSGISGSCFAPNPVLLISAPNKINFTLRKKRRLVSAVSYPEFASALE 60
DB 1 MALSMFSGISGSCFAPNPVLLISAPNKINFTLRKKRRLVSAVSYPEFASALE 60
QY 61 ETRKRIVLEPSHLOEKYSMTGLDQKTELQMLAFKSSKIRLLRSMAIENETMQVDFAGF 120
DB 61 ETRKRIVLEPSHLOEKYSMTGLDQKTELQMLAFKSSKIRLLRSMAIENETMQVDFAGF 120
QY 121 MEPEYDPIFCANFTSTNNVIVLDLNPRLHQLDQDTYODKYKIMSIYHKYAEETPPW 180
DB 121 MEPEYDPIFCANFTSTNNVIVLDLNPRLHQLDQDTYODKYKIMSIYHKYAEETPPW 180
QY 181 GSKLTGESIKFSPVLMWTRFSSSEKHKALFSAFLEYQOAWLEMTIOVREMEPSHYRA 240
DB 181 GSKLTGESIKFSPVLMWTRFSSSEKHKALFSAFLEYQOAWLEMTIOVREMEPSHYRA 240
QY 241 NCEAOHKYLTWRAQKDPGHGLLKRVLGSAKAKELLRDLFNGVDELGKTFIDYPEPYOT 300
DB 241 NCEAOHKYLTWRAQKDPGHGLLKRVLGSAKAKELLRDLFNGVDELGKTFIDYPEPYOT 300

```

```

DB 241 NCEAOHKYLTWRAQKDPGHGLLKRVLGSAKAKELLRDLFNGVDELGKTFIDYPEPYOT 300
QY 301 EDGTVSDKRSITIGKSYETRPMDLTGQFTG 329
DB 301 EDGTVSDKRSITIGKSYETRPMDLTGQFTG 329

```

RESULT 6

```

US-09-870-406a-55
; Sequence 55, Application US/09870406A
; Publication No. US20030104379A1
; GENERAL INFORMATION:
; APPLICANT: LAGARIAS, JOHN
; APPLICANT: KOICHI, TAKAYUKI
; APPLICANT: FRANKENBERG, NICOLE
; APPLICANT: GAMBETTA, GREGORY
; APPLICANT: MONTGOMERY, BERONDA
; TITLE OF INVENTION: HY2 FAMILY OF BILLIN REDUCTASES
; FILE REFERENCE: 407T-907720US
; CURRENT APPLICATION NUMBER: US/09/870,406A
; PRIOR FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: 60/271,758
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 60/210,286
; PRIOR FILING DATE: 2000-06-08
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 55
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Arapidopsis thaliana
US-09-870-406a-55

```

Query Match

```

Best Local Similarity 100.0%; Score 1724; DB 9; Length 329;
Pred. No. 3,4e-158;
Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MALSMFSGISGSCFAPNPVLLISAPNKINFTLRKKRRLVSAVSYPEFASALE 60
DB 1 MALSMFSGISGSCFAPNPVLLISAPNKINFTLRKKRRLVSAVSYPEFASALE 60
QY 61 ETRKRIVLEPSHLOEKYSMTGLDQKTELQMLAFKSSKIRLLRSMAIENETMQVDFAGF 120
DB 61 ETRKRIVLEPSHLOEKYSMTGLDQKTELQMLAFKSSKIRLLRSMAIENETMQVDFAGF 120
QY 121 MEPEYDPIFCANFTSTNNVIVLDLNPRLHQLDQDTYODKYKIMSIYHKYAEETPPW 180
DB 121 MEPEYDPIFCANFTSTNNVIVLDLNPRLHQLDQDTYODKYKIMSIYHKYAEETPPW 180
QY 181 GSKLTGESIKFSPVLMWTRFSSSEKHKALFSAFLEYQOAWLEMTIOVREMEPSHYRA 240
DB 181 GSKLTGESIKFSPVLMWTRFSSSEKHKALFSAFLEYQOAWLEMTIOVREMEPSHYRA 240
QY 241 NCEAOHKYLTWRAQKDPGHGLLKRVLGSAKAKELLRDLFNGVDELGKTFIDYPEPYOT 300
DB 241 NCEAOHKYLTWRAQKDPGHGLLKRVLGSAKAKELLRDLFNGVDELGKTFIDYPEPYOT 300
QY 301 EDGTVSDKRSITIGKSYETRPMDLTGQFTG 329
DB 301 EDGTVSDKRSITIGKSYETRPMDLTGQFTG 329

```

RESULT 7

```

US-10-159-901-54
; Sequence 54, Application US/10159901
; Publication No. US20030073235A1
; GENERAL INFORMATION:
; APPLICANT: LAGARIAS, JOHN
; APPLICANT: KOICHI, TAKAYUKI
; APPLICANT: FRANKENBERG, NICOLE
; APPLICANT: GAMBETTA, GREGORY
; APPLICANT: MONTGOMERY, BERONDA

```

;; TITLE OF INVENTION: LIGHT CONTROLLED GENE EXPRESSION UTILIZING HETEROLOGOUS PHYTOCHROME
;; FILE REFERENCE: 407T-907731US
;; CURRENT APPLICATION NUMBER: US/10/159,901
;; CURRENT FILING DATE: 2002-05-29
;; PRIOR APPLICATION NUMBER: 60/294,463
;; PRIOR FILING DATE: 2001-05-29
;; NUMBER OF SEQ ID NOS: 57
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 54
;; LENGTH: 255
;; TYPE: PRF
;; ORGANISM: No. US20030073235A1toc punctiliforme
US-10-159-901-54

Query Match 19.7%; Score 339.5; DB 9; Length 255;
Best Local Similarity 31.1%; Pred. No. 1.1e-24;
Matches 76; Conservative 51; Mismatches 108; Indels 9; Gaps 3;

QY 45 RVSAVSYKEPAESALAEETRRKRIYLPESHLOEKYSMTGDLGKTELQ-----MLAFKSSK 98
DB 5 RSDVTLXQPLFDYAIAYMRSRLDLEPYPIPTGFESNAVYKGNQGEVVTTSYAFQFAK 64
QY 99 IRLRSMAIE-NETMOVDFRAGMEPEYDPIFCANFSTNNIVVLDLPHQLTDQT 157
DB 65 LKQIRAAHVQGNLSQVLYNFIPLPHNLVDLPEFGADLVTLPGGHLIALDQPLFR--DDS 122
QY 158 DYODKYXNKIMSIYHKAETFPWCKLTGESIKFSPSLVMWTRSSSEKHKALFSAFLE 217
DB 123 AYQAKTYEPLIFPIFHAHQHLSWGDPEEAQPFPSAFMLTRQETAVVETQVFAAFKD 182
QY 218 YQAMLEMTIQVREMEPSHVANCEAOKHYLTWRQKDPGHGLKRLVGEAKAKELLRD 277
DB 183 YLKAAYLDFVQAEVATDSQNLVAIKQQLRYLRRAEKDPAKMFKRYGAEMTEEYIHG 242
QY 278 FLFN 281
DB 243 FLFD 246

RESULT 8
US-09-870-406a-54
;; Sequence 54, Application US/09870406A
;; Publication No. US20030104379A1
;; GENERAL INFORMATION:
;; APPLICANT: LAGARIAS, JOHN
;; APPLICANT: KOICHI, TAKAYUKI
;; APPLICANT: FRANKENBERG, NICOLE
;; APPLICANT: GAMBETTA, GREGORY
;; APPLICANT: MONTGOMERY, BERONDA
;; TITLE OF INVENTION: HY2 FAMILY OF BILIN REDUCTASES
;; FILE REFERENCE: 407T-907720US
;; CURRENT APPLICATION NUMBER: US/09/870,406A
;; CURRENT FILING DATE: 2002-09-04
;; PRIOR APPLICATION NUMBER: 60/271,758
;; PRIOR FILING DATE: 2001-02-26
;; PRIOR APPLICATION NUMBER: 60/210,286
;; PRIOR FILING DATE: 2000-06-08
;; NUMBER OF SEQ ID NOS: 57
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 54
;; LENGTH: 255
;; TYPE: PRF
;; ORGANISM: No. US20030104379A1toc punctiliforme
US-09-870-406a-54

Query Match 19.7%; Score 339.5; DB 9; Length 255;
Best Local Similarity 31.1%; Pred. No. 1.1e-24;
Matches 76; Conservative 51; Mismatches 108; Indels 9; Gaps 3;

QY 45 RVSAVSYKEPAESALAEETRRKRIYLPESHLOEKYSMTGDLGKTELQ-----MLAFKSSK 98
DB 5 RSDVTLXQPLFDYAIAYMRSRLDLEPYPIPTGFESNAVYKGNQGEVVTTSYAFQFAK 64

QY 99 IRLRSMAIE-NETMOVDFRAGMEPEYDPIFCANFSTNNIVVLDLPHQLTDQT 157
DB 65 LKQIRAAHVQGNLSQVLYNFIPLPHNLVDLPEFGADLVTLPGGHLIALDQPLFR--DDS 122
QY 158 DYODKYXNKIMSIYHKAETFPWCKLTGESIKFSPSLVMWTRSSSEKHKALFSAFLE 217
DB 123 AYQAKTYEPLIFPIFHAHQHLSWGDPEEAQPFPSAFMLTRQETAVVETQVFAAFKD 182
QY 218 YQAMLEMTIQVREMEPSHVANCEAOKHYLTWRQKDPGHGLKRLVGEAKAKELLRD 277
DB 183 YLKAAYLDFVQAEVATDSQNLVAIKQQLRYLRRAEKDPAKMFKRYGAEMTEEYIHG 242
QY 278 FLFN 281
DB 243 FLFD 246

RESULT 9
US-10-159-901-38
;; Sequence 38, Application US/10159901
;; Publication No. US20030073235A1
;; GENERAL INFORMATION:
;; APPLICANT: LAGARIAS, JOHN
;; APPLICANT: KOICHI, TAKAYUKI
;; APPLICANT: FRANKENBERG, NICOLE
;; APPLICANT: GAMBETTA, GREGORY
;; APPLICANT: MONTGOMERY, BERONDA
;; TITLE OF INVENTION: LIGHT CONTROLLED GENE EXPRESSION UTILIZING HETEROLOGOUS PHYTOCHROME
;; FILE REFERENCE: 407T-907731US
;; CURRENT APPLICATION NUMBER: US/10/159,901
;; CURRENT FILING DATE: 2002-05-29
;; PRIOR APPLICATION NUMBER: 60/294,463
;; PRIOR FILING DATE: 2001-05-29
;; NUMBER OF SEQ ID NOS: 57
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 38
;; LENGTH: 257
;; TYPE: PRF
;; ORGANISM: PROCHLOROCCUS MARINUS
US-10-159-901-38

Query Match 12.5%; Score 215; DB 9; Length 257;
Best Local Similarity 26.1%; Pred. No. 1.1e-12;
Matches 59; Conservative 48; Mismatches 103; Indels 16; Gaps 6;

QY 68 LPSHQLQEKYSMTGDLK-----TELQMLAFKSSKIRLRSMAIE-NETMOVDFRAG 119
DB 35 LEPYVSHDLSESTIGSRNPVHTTLTMAA-KFEKIKOYRLACIKGSELSVFNLI 93
QY 120 FMEPEYDPIFCANFSTNNIVVLDLPHQLTDQTDYODKYXNKIMSIYHKAETFP 179
DB 94 HPLNDVLDLPEFGADLVTLNGHLIALDQPALKL--DNHTENWPRRLPLDHDQSLLP 151
QY 180 WCKLTGESIKFSPSLVMWTRSSSEK---HKALFSAFLEYQAMLEMTIQVREMEP 235
DB 152 SGGEIPKEAPEYFSPFLMSRLPSKESDNIISEILRPAGGSLIYEL-LHIKAPLKK 210
QY 236 SHVRANCEAOKHYLTWRQKDPGHGLKRLVGEAKAKELLRDPLFN 281
DB 211 ERAKLITBQAKAYINRSTKIDPARAMLCRFYKEMTEYIHKVLFN 256

RESULT 10
US-10-159-901-52
;; Sequence 52, Application US/10159901
;; Publication No. US20030073235A1
;; GENERAL INFORMATION:
;; APPLICANT: LAGARIAS, JOHN
;; APPLICANT: KOICHI, TAKAYUKI
;; APPLICANT: FRANKENBERG, NICOLE
;; APPLICANT: GAMBETTA, GREGORY
;; APPLICANT: MONTGOMERY, BERONDA
;; TITLE OF INVENTION: LIGHT CONTROLLED GENE EXPRESSION UTILIZING HETEROLOGOUS PHYTOCHROME


```
FILE REFERENCE: 407T-907731US
CURRENT APPLICATION NUMBER: US/10/159,901
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: 60/294,463
PRIOR FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 57
SOFTWARE: PatentIn version 3.0
SEQ ID NO 52
LENGTH: 257
TYPE: PRT
ORGANISM: Prochlorococcus marinus
US-10-159-901-52
```

```
Query Match          12.5%: Score 215; DB 9; Length 257;
Best Local Similarity 26.1%: Pred. No. 1.1e-12;
Matches 59; Conservative 48; Mismatches 103; Indels 16; Gaps 6;
```

```
QY 68 LEPSHLOEKYSMTGLDCK-----TELQMLAFKSSKIRLLRSMAT-NEFMQVDFPAG 119
DB 35 IEPPVSHDFLSKESITGSRNPNVHTLTWAA-KFEKIKQVRLACIGGSLSVFNLLI 93
QY 120 FMEBYDPTFCANFEFTSTNNIVLDLNPLOLTDQDYODKYNNKMSIYHKYAEFFP 179
DB 94 HPLNDYDLPFGADPVTLPNGHLALDQPALKL--DNIRHNWNPRLIPLDHMQSLLP 151
QY 180 WGRKLTGESIKFFSPPLVWMTFRSSSKK-----HKALFSAPLEYQAWLEMTIQVREMEP 235
DB 152 SGGEIPKAEPEYFSPGFLMSRLPLSKESDNIISLIRAFGEYLSLYTEL-LHIKAPLKK 210
QY 236 SHVRANCEAOKHYLTWRQKDPGHGLKRLVGEAKAKELDLFLN 281
DB 211 ERAKLIEGOKAYINVRSTKDPARAMLCRFYGKEMTEDEYIHKVLFN 256
```

```
RESULT 11
US-09-870-406a-38
Sequence 38, Application US/09870406A
Publication No. US20030104379A1
GENERAL INFORMATION:
APPLICANT: LAGARIAS, JOHN
APPLICANT: KOICHI, TAKAYUKI
APPLICANT: FRANKENBERG, NICOLE
APPLICANT: GAMBETTA, GREGORY
APPLICANT: MONTGOMERY, BERONDA
TITLE OF INVENTION: HY2 FAMILY OF BILIN REDUCTASES
FILE REFERENCE: 407T-907720US
CURRENT APPLICATION NUMBER: US/09/870,406A
CURRENT FILING DATE: 2002-09-04
PRIOR APPLICATION NUMBER: 60/271,758
PRIOR FILING DATE: 2001-02-26
PRIOR APPLICATION NUMBER: 60/210,286
PRIOR FILING DATE: 2000-06-08
NUMBER OF SEQ ID NOS: 57
SOFTWARE: PatentIn version 3.0
SEQ ID NO 38
LENGTH: 257
TYPE: PRT
ORGANISM: PROCHLOROCOCCUS MARINUS
US-09-870-406a-38
```

```
Query Match          12.5%: Score 215; DB 9; Length 257;
Best Local Similarity 26.1%: Pred. No. 1.1e-12;
Matches 59; Conservative 48; Mismatches 103; Indels 16; Gaps 6;
```

```
DB 152 SGGEIPKAEPEYFSPGFLMSRLPLSKESDNIISLIRAFGEYLSLYTEL-LHIKAPLKK 210
QY 236 SHVRANCEAOKHYLTWRQKDPGHGLKRLVGEAKAKELDLFLN 281
DB 211 ERAKLIEGOKAYINVRSTKDPARAMLCRFYGKEMTEDEYIHKVLFN 256
```

```
RESULT 12
US-09-870-406a-52
Sequence 52, Application US/09870406A
Publication No. US20030104379A1
GENERAL INFORMATION:
APPLICANT: LAGARIAS, JOHN
APPLICANT: KOICHI, TAKAYUKI
APPLICANT: FRANKENBERG, NICOLE
APPLICANT: GAMBETTA, GREGORY
APPLICANT: MONTGOMERY, BERONDA
TITLE OF INVENTION: HY2 FAMILY OF BILIN REDUCTASES
FILE REFERENCE: 407T-907720US
CURRENT APPLICATION NUMBER: US/09/870,406A
CURRENT FILING DATE: 2002-09-04
PRIOR APPLICATION NUMBER: 60/271,758
PRIOR FILING DATE: 2001-02-26
PRIOR APPLICATION NUMBER: 60/210,286
PRIOR FILING DATE: 2000-06-08
NUMBER OF SEQ ID NOS: 57
SEQ ID NO 52
LENGTH: 257
TYPE: PRT
ORGANISM: Prochlorococcus marinus
US-09-870-406a-52
```

```
Query Match          12.5%: Score 215; DB 9; Length 257;
Best Local Similarity 26.1%: Pred. No. 1.1e-12;
Matches 59; Conservative 48; Mismatches 103; Indels 16; Gaps 6;
```

```
QY 68 LEPSHLOEKYSMTGLDCK-----TELQMLAFKSSKIRLLRSMAT-NEFMQVDFPAG 119
DB 35 IEPPVSHDFLSKESITGSRNPNVHTLTWAA-KFEKIKQVRLACIGGSLSVFNLLI 93
QY 120 FMEBYDPTFCANFEFTSTNNIVLDLNPLOLTDQDYODKYNNKMSIYHKYAEFFP 179
DB 94 HPLNDYDLPFGADPVTLPNGHLALDQPALKL--DNIRHNWNPRLIPLDHMQSLLP 151
QY 180 WGRKLTGESIKFFSPPLVWMTFRSSSKK-----HKALFSAPLEYQAWLEMTIQVREMEP 235
DB 152 SGGEIPKAEPEYFSPGFLMSRLPLSKESDNIISLIRAFGEYLSLYTEL-LHIKAPLKK 210
QY 236 SHVRANCEAOKHYLTWRQKDPGHGLKRLVGEAKAKELDLFLN 281
DB 211 ERAKLIEGOKAYINVRSTKDPARAMLCRFYGKEMTEDEYIHKVLFN 256
```

```
RESULT 13
US-10-159-901-36
Sequence 36, Application US/10159901
Publication No. US20030073235A1
GENERAL INFORMATION:
APPLICANT: LAGARIAS, JOHN
APPLICANT: KOICHI, TAKAYUKI
APPLICANT: FRANKENBERG, NICOLE
APPLICANT: GAMBETTA, GREGORY
APPLICANT: MONTGOMERY, BERONDA
TITLE OF INVENTION: LIGHT CONTROLLED GENE EXPRESSION UTILIZING HETEROLOGOUS PHYTOC
FILE REFERENCE: 407T-907731US
CURRENT APPLICATION NUMBER: US/10/159,901
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: 60/294,463
PRIOR FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 57
SOFTWARE: PatentIn version 3.0
```

```

; SEQ ID NO 36
;
; LENGTH: 257
;
; TYPE: prt
;
; ORGANISM: Synechococcus sp
;
US-10-159-901-36

```

Query Match	12.2%	Score 210.5	DB 9	Length 257
Best Local Similarity	24.3%	Pred. No. 3e-12		
Matches 63	Conservative 55	Mismatches 126	Indels 15	Gaps 8

QY	34	TLRRRRKKFFLLRVASVSKFEAESLLETKRKRIYVEPHLOEKT---	SSMTGLDCKT---	87
Db	2	TNÖRFKSDPYDNIGSWMQPFLDEAIKRL-EGIANEPFPVDRFLQÖRDQIGCSKKSIFV		60
QY	88	ELÖMLAEKSSKTRILIRSMAlE-NETMÖYDFPAGEMEPEYDPIFCANFFSTNNIVIID		146
Db	61	TTATWACKTEKTEFRÖVRAACVAGSAAVSLNFIINFKSTYGGJPFPGDGLVTFPAGHILMD		120
QY	147	LNPLHOLDÖDYÖDKYKKNLMSIYHKYAEFPWCGKLTGESIKFESPLVMTRRSSKE		206
Db	121	LÖPAIK-IDEV-HITTHVMDRLPIPERKRDÖLPYGGPIPEBÖQPFSSGFLMTRLPJGE		178
QY	207	KHKALFS---AFLEYQAMLEMTLOVBEEMEPSHVRANCEÖHKYLLTRRAÖKDGHILL		262
Db	179	GDELQSGIVRAFNNDYIDLTYLEIAASA-ERVDERSEVILLQÖQRXYTDVRAEKDPAQWIL		237
QY	263	KRLVGEAKAKELLDNDFLN	281	
Db	238	TRFHGSEWTEAVIHTVLFD	256	

```

RESULT 14
US-10-159-901-50
: Sequence 50, Application US/10159901
: Publication NO. US20030073235A1
: GENERAL INFORMATION:
: APPLICANT: LAGARIAS, JOHN
: APPLICANT: KOICHI, TAKAYUKI
: APPLICANT: FRANKENBERG, NICOLE
: APPLICANT: GAMBETTA, GREGORY
: APPLICANT: MONTGOMERY, BERNONDA
: TITLE OF INVENTION: LIGHT CONTROLLED GENE EXPRESSION UTILIZING HETEROLOGOUS PHYTOCHROME
: FILE REFERENCE: 407T-907731US
: CURRENT APPLICATION NUMBER: US/10/159,901
: CURRENT FILING DATE: 2002-05-29
: PRIOR APPLICATION NUMBER: 60/294,463
: PRIOR FILING DATE: 2001-05-29
: NUMBER OF SEQ ID NOS: 57
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 50
: LENGTH: 257
: TYPE: PRN
: ORGANISM: Synechocystis sp.
US-10-159-901-50

```

Query Match	12.2%	Score 210.5	DB 9	length 257
Best Local Similarity	24.3%	Pred. No. 3e-12		
Matches	63	Conservative	55	Mismatches 126; Indels 15; Gaps 8

QY	34	TLRRKRKFFLARSVSKYKFAESHLEETRRKIYLEPRLQKY----	87
Db	2	TNQPKSTDPVNIQSGMSQPFLEDAIKKL--ELNAYEPVPVPRLOREDQTSKSKSI	60
QY	88	ELQMLAEKSKIRLLRSMIAIE--NETMOYDFAGEMEPEYDPIFCANFTSTNVNI	146
Db	61	TTATWACKTEKFRQVRAACVSAGSAVASVINEFINRSTYGLPFFEGDILVTPFAGH	120
QY	147	LNPIHOLDQDYODKYKIMSIYHNKYAEFFPMGCKLTGESIKFESPLVMTTRSS	206
Db	121	LQPAIK--TDEV--HTTHWMDRLPIPERNRDQLPYGRIPREARQPFPSGFLTR	178
QY	207	KHKALFS----AFLEYQAMLEMTIQVADEEMPSHYRANCEAOKHLLTVRAKQD	262

Db 179 GDEIGSIYRPAINDLDLLELAASA-ERYIDENSEVLLGQGRKTYTDYRAEKDPARGML 237

QY 263 KRLVGEAKAKELLRDLEFN 281

Db 238 TREHGSSEMTPEAYIHTVLF 256

RESULT 15
 US-09-870-406A-36
 Sequence 36, Application US/09870406A
 Publication No. US20030104379A1
 GENERAL INFORMATION:
 APPLICANT: LAGARIAS, JOHN
 APPLICANT: KOICHI, TAKAYUKI
 APPLICANT: FRANKENBERG, NICOLE
 APPLICANT: GAMBETTA, GREGORY
 APPLICANT: MONTGOMERY, BERONDA
 TITLE OF INVENTION: HY2 FAMILY OF BILIN REDUCTASES
 FILE REFERENCE: 4071-907720US
 CURRENT APPLICATION NUMBER: US/09/870,406A
 CURRENT FILING DATE: 2002-09-04
 PRIOR APPLICATION NUMBER: 60/271,758
 PRIOR FILING DATE: 2001-02-26
 PRIOR APPLICATION NUMBER: 60/210,286
 PRIOR FILING DATE: 2000-06-08
 NUMBER OF SEQ ID NOS: 57
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 36
 LENGTH: 257
 TYPE: PR1
 ORGANISM: *Synechococcus* sp.
 US-09-870-406A-36

[illegible]

Search completed: June 24, 2003, 18:46:50
Job time : 161 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 18:13:35 ; Search time 27 Seconds

1171.416 Million cell updates/sec

Title: US-09-870-406A-33

Sequence: 1 MALSMEEFGFSIGSCFKAPNP.....SIIGKSYETRPWDLTGQFIG 329

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
Database : PIR_73:*
1: plr1:*
2: plr2:*
3: plr3:*
4: plr4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	210.5	12.2	257	2	B46448	hypothetical protease
2	158	9.2	236	2	S31058	hypothetical protease
3	157	9.1	256	2	A46448	orf36 3' of cpab
4	121.5	7.0	245	2	AD2269	hypothetical protease
5	110.5	6.4	549	1	A54541	alpha-amylase (EC
6	105.5	6.1	248	2	S76709	hypothetical protease
7	105	6.1	548	1	ALBSF	alpha-amylase (EC
8	104.5	6.1	549	1	A24549	alpha-amylase (EC
9	102.5	5.9	311	2	T30905	alpha-amylase (EC
10	95.5	5.5	595	2	B97866	hypothetical protease
11	95.5	5.5	1125	1	F70177	DNA primase (EC 2.
12	95	5.5	574	2	B35149	transcription-repa
13	94.5	5.5	649	2	T33741	ipap protein - Shl
14	94	5.5	1452	2	A97323	DNA-binding protease
15	93.5	5.4	374	2	H70415	DNA-polymerase III
16	93.5	5.4	386	2	G8652	conserved hypot
17	93.5	5.4	2819	2	A90551	GTP-binding protease
18	91.5	5.3	841	2	B71212	conserved hypot
19	91.5	5.3	3119	2	T18414	hypothetical protease
20	91	5.3	1073	2	S69079	protein g377 - mal
21	91	5.3	2091	2	A97077	hypothetical protease
22	91	5.3	2100	2	T38128	hypothetical protease
23	90.5	5.2	1046	2	A86790	t7123.15 protein -
24	90.5	5.2	2034	2	T22147	ATP-dependent dsDN
25	90	5.2	392	2	B96554	hypothetical protease
26	90	5.2	553	2	G71892	unknown protein, 1
27	90	5.2	817	2	S53921	probable osmoprote
28	89.5	5.2	1120	2	S67208	hypothetical protease
29	89.5	5.2	1275	2	AD0332	hypothetical protease

30	89	5.2	1045	2	H82934	type I restriction
31	89	5.2	1254	2	T24897	hypothetical prote
32	89	5.2	1451	2	D64203	DNA polymerase III
33	89	5.2	4450	2	JX0340	gramicidin S synth
34	88.5	5.1	433	2	A29266	apolipoprotein B -
35	88.5	5.1	549	1	A24436	alpha-amyase (EC
36	88.5	5.1	711	2	S66749	hypothetical prote
37	88.5	5.1	936	2	H71728	2-oxoglutarate deh
38	88.5	5.1	1827	2	T16270	hypothetical prote
39	88	5.1	518	2	D84977	glutamate-cysteine
40	88	5.1	537	2	B90598	ABC transporter at
41	88	5.1	549	2	A31584	carboxylesterase (
42	88	5.1	540	2	AX0054	carboxylesterase (
43	88	5.1	693	2	A10966	4-alpha-glucanotra
44	87.5	5.1	602	2	E90568	DNA primase (impor
45	87.5	5.1	1352	2	G71051	probable ATP-depen

ALIGNMENTS

RESULT 1
B46448

hypothetical protein (rpcb 5' region) - *Synechococcus* sp.

C;Species: Synechococcus sp.

C:/Accession: B46448: S31066: S31059
C:/Date: 10-Jun-1993 #sequence_revision 18-NOV-1994 #ccalc_change 00 OCT 1995

R;Wilbanks, S.M.; Glazer, A.N.

J. Biol. Chem. 268, 1226-1235, 1993

chococcus sp. WH8020.

A;Reference number: A45045; MUID:93123238; PMID:84193255

A; Accession: B46448

A: Molecule type: DNA
A: status: preliminary; not compared with conceptual

A;Residues: 1-257 <

A; Cross-references: EMBL:M95288; NID:g154551; PIDN:AAA27344.1;

A; Experimental source

R;de Lormier, R.; Wilbanks, S.M.; Glazer, A.N.

Plant Mol. Biol. 21, 225-237,

A;Title: Genes of the R-phycocyanin II locus of marine *Synechococcus* sp. PCC 7002. A:Reference number: S31066: MTID:93144698: BMTD:8425055

A/Accession: S31066

A; molecule type: DNA

A;Residues: 231-257
A;Cross-references: EMBL: M95288

RECEIVED

Query Match

Best Local Similar

machines 03; conservative

QY 34

[illegible][illegible]

RESULT 2

S31058

hypothetical protein (cpeb 3' region) - *Synechococcus* sp. (strain WH8020)C:Species: *Synechococcus* sp.

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 08-Oct-1999

C:Accession: S31058

R:Wilbanks, S.M.; Glazer, A.N.

submitted to the EMBL data library, October 1992

A:Reference number: S31047

A:Accession: S31058

A:Molecule type: DNA

A:Residues: 1-236 <WIL>

A:Cross-references: EMBL:M95288; NID:g154551; PIDN:AAA27343.1; PID:g154563

Query Match

Best Local Similarity 9.2%; Score 158; DB 2; Length 236;

Matches 60; Conservative 41; Mismatches 89; Indels 56; Gaps 11;

QY 71 SHLOEKYSMT-----GLOGKTELOMLAKSSKI-----RLLRSM 105

DB 4 SFLNELHSDITKRGSGPLPEGL-----ECRSSKSSSVIOSWLMVPGFRMRVTRLD 58

QY 106 ALENETMOVDFAGMEPEYDPIPCAN--FFTSTNNIVYLDLNLPHQLDQDQYDQY 163

DB 59 A--GDSLQVENSVAVPDYNDHPDLMGVDLLMGAROKIVAVLDQPLVQ---DKDYLDY 113

QY 164 YKIKMSIYHKYAEFPWGGKLTG--ESIKFSP---LVNMTFRSSSEKHKAL-----FSA 214

DB 114 FSGLELMORFPD-----LNGEETMRSPDNOYFSSWMLFCRGAQADLSLPKAFSA 166

QY 215 FLEYQAVMLMTIQVREMEPSHVRANCEAQHKYLTWRAQKDPGGLKRLVGEAKAKEL 274

DB 167 FLKAYWDLHDNAKSPSTIPPEYK---NLQDKYDIYAEKDPADHGLTSHFGKDWNSRF 223

QY 275 LRDELTF 280

DB 224 LHEFLF 229

RESULT 3

A46448

orf236 3' of cpeb - *Synechococcus* sp.C:Species: *Synechococcus* sp.

C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 23-Mar-1995

C:Accession: A46448

R:Wilbanks, S.M.; Glazer, A.N.

J. Biol. Chem. 268, 1226-1235, 1993

A:Title: Rod structure of a phycoerythrin II-containing phycobilisome. I. Organization of *Synechococcus* sp. WH8020

A:Reference number: A45045; MUID:93123238; PMID:8419325

A:Accession: A46448

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-256 <WIL>

A:Experimental source: WH8020

A:Note: sequence extracted from NCBI backbone (NCBIP:121984)

Query Match

Best Local Similarity 9.1%; Score 157; DB 2; Length 256;

Matches 47; Conservative 34; Mismatches 77; Indels 24; Gaps 7;

QY 110 ETMOVDFAGMEPEYDPIPCAN--FFTSTNNIVYLDLNLPHQLDQDQYDQY 167

DB 81 DSIQVENSVAVPDYNDHPDLMGVDLLMGAROKIVAVLDQPLVQ---DKDYLDYFSG 137

QY 168 MSYHYAETFPWGGKLTG--ESIKFSP---LVNMTFRSSSEKHKAL-----FSAFLEY 218

DB 138 KEINQFPD-----LNGEETMRSPDNOYFSSWMLFCRGAQADLSLPKAFSAFKA 190

QY 219 YQAMLEMTIQVREMEPSHVRANCEAQHKYLTWRAQKDPGGLKRLVGEAKAKEL 278

DB 191 YMDLHDNAKSPSTIPPEYK---NLQDKYDIYAEKDPADHGLTSHFGKDWNSRF 247

QY 279 LF 280

DB 248 LF 249

RESULT 4

AD2269

hypothetical protein alr3707 [imported] - *Nostoc* sp. (strain PCC 7120)C:Species: *Nostoc* sp.A:Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C:Accession: AD2269

R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kunitz, T.; Sasamoto, S.; Watanabe, A.; Irigu

Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata

DNA Res. 6, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AD2269

A:Molecule type: DNA

A>Status: preliminary

A:Residues: 1-245 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA075406.1; PID:g17132841; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr3707

Query Match

Best Local Similarity 7.0%; Score 121.5; DB 2; Length 245;

Matches 52; Conservative 49; Mismatches 113; Indels 27; Gaps 10;

QY 52 KEAESALEETRRKRIYERSHLOEKYSMTG--LDG--KTELOMLAFSSKIRLLR--SMAE 108

DB 18 RQADCIIEYVHQHLDLSYHLPAELGYEGRLGEGKLIENRCYOTPOFRMHLELAKV 77

QY 109 NETMOVDFAGMEPEYDPIPCAN--FFTSTNNIVYLDLNLPHQLDQDQYDQY 167

DB 78 GNMULDILHCVMFPRPYDLPFGCDLVGGRGOISAIDLSPVH--LDRT--LPESYNSAL 134

QY 168 MSI---YHKYAEFPWGGKLTGSEIKFSP---LVNMTFRSSSEKHKALFSAFLEYQAM 223

DB 135 TSLNTLNFSGPRLPEMG-----NIFSDCIVFRSSPEE-----AMFLGVAREPL 181

QY 224 EMTIQVREMEPSHVRANCE---AQHKYLTWRAQKDPGGLKRLVGEAKAKELRDLF 280

DB 182 QVHCQALIASPVSAEOKQIILAGOHNYCSKQKQDKTRRYLEKAFGVDMENVTYVLF 241

QY 281 N 281

DB 242 D 242

RESULT 5

A54541

alpha-amyase (EC 3.2.1.1) precursor - *Bacillus stearothermophilus* (strain DNI792)

N:Alternate names: 1,4-alpha-D-glucan glucanohydrolase

C:Species: *Bacillus stearothermophilus*

C:Date: 28-Oct-1994 #sequence_revision 18-Aug-1995 #text_change 13-Jun-1997

C:Accession: A54541

R:Jorgensen, P.L.; Poulsen, G.B.; Diderichsen, B.

FEBS Microbiol. Lett. 77, 271-276, 1991

A:Title: Cloning of a chromosomal alpha-amyase gene from *Bacillus stearothermophilus*

A:Reference number: A54541

A:Accession: A54541

A:Molecule type: DNA

A:Residues: 1-549 <JOR>

A:Cross-references: GB:X59476

A:Experimental source: chromosomal DNA of strain DNI792

C:Comment: Alpha-amyase genes have been found on plasmids and in multiple copies on

A:Start codon: GTG

A:Function: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds

A:Pathway: glycogen/starch degradation

C:Superfamily: alpha-amylase, amyloliquefaciens type; alpha-amylase core homology
 C:Keywords: extracellular protein; glycosidase; heat-stable protein; hydrolase; polysac
 F:1-34/Domain: signal sequence #status predicted <SIG>
 F:35-549/Product: alpha-amylase #status predicted <AMT>
 F:235-368/Domain: alpha-amylase core homology <AMT>
 F:139,237,272/Binding site: calcium (Asp, Asp, His) #status predicted
 F:268,298,365/Active site: Asp, Glu, Asp #status predicted

Query Match 6.1%; Score 110.5; DB 1; Length 549;
 Best Local Similarity 19.4%; Pred. No. 0.44;
 Matches 68; Conservative 50; Mismatches 120; Indels 113; Gaps 16;

QY 50 SYKEFAESALEETRRRIYLPESHLOEKYSSMTGLDGKTELQMLAKSSKIRLLRMAIEN 109
 DB 229 NVDYLMVADLDHDEHVEVTELKNMGKWTNTNIDG-----FRIDAKYH 272
 QY 110 ETMOVF-DPAGFMEPEYDPIFCANFTSTNNIVLNLPLHQLDQTD-----Y 159
 DB 273 IKFSEFPDMLSVRSQGTGKPLETVEYWS-----YDINKLHNYITRTDGTMSLFDAPL 325
 QY 160 QDKYV-----NKMSIYHKYAEF-----PMGSKLTG 186
 DB 326 HNKFTASKSGGAFDMRTLMTNTLMKDDPTLAVTFVNDHDEPQALQSWDPM----- 379
 QY 187 ESIRKFSPLVMTTRFSSSEKHKALFSAFLEYQAWLEMTIQVREMEPSHVANCEA-- 244
 DB 380 -----FKPLA-YAFILTRQEGPCVF--YGDYIGIPQYNIPSLKSKIDPLIARDYAG 431
 QY 245 -QHKYLT-----WRAQ---KDPGHGLKRL-----VGEAKAKELLRDPLFNGV 283
 DB 432 TQHDYLDHSDIIGMTREGTEKPGSGLAALITDGGSKMMVYKQHGAKVYFDLTGNRS 491
 QY 284 DELGKRTFIDFPEYQTEDEGTVS-----DKRSIIIGKSYETRRPMDLTGQFI 328
 DB 492 DTVTINS--DGMGEKNGSVSVWPKRTYVTIARITTRPW--TGSEFV 538

RESULT 6

S76709
 hypothetical protein - Synechocystis sp. (strain PCC 6803)
 C:Species: Synechocystis sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
 C:Accession: S76709
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.;
 O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
 S.

A:Reference number: S74322; MUID:97061201; PMID:8905231
 A:Accession: S76709
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-248 <KAN>
 A:Cross-references: EMBL:D64004; GB:AB001339; NID:g1001701; PIDN:BA010653.1; PID:d101130
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 6.1%; Score 105.5; DB 2; Length 248;
 Best Local Similarity 21.4%; Pred. No. 0.39;
 Matches 50; Conservative 49; Mismatches 110; Indels 25; Gaps 10;

QY 58 ALEETRRRIYLPESHLOEKYSSMTG-LDG-KTELQMLAFKSKIRLLR-SMAIENETMOV 114
 DB 27 AIAASWQSLPLKPYOLPEDLGVEGRLEGKLVITNRCYQTFQPKMHLAKVKGGLDI 86
 QY 115 FDFAGFMEPEYDPIFCANFTSTNNIVLNLPLHQLDQTDQD--DKYYNKIMS-- 169
 DB 87 LHCWAFPEPIYGLPLFGCDIVAGPGVSAIADISP-----TQSDROLPAVYQKSLAELG 141
 QY 170 --IYKVAETFPWGKLGESIKFSPVLMWTRFSSSEKHKALFSAFLEYQAWLEMTI 227
 DB 142 QPEFQOQRELPWVG-----EIFSEYCLFIRPSNVTSEER-FVQRYVDFLQIHOSI 192

QY 228 QVREMEPSHVANCEAOKHYLTWPAOKDPGHLKRLVGEAKAKELLRDPLEN 281
 DB 193 -VAEFLSEAQTLERHQGLIHQOQOQKNDKTRVLEKAFGEAMERYSQVLF 245

RESULT 7

ABSF
 alpha-amylase (EC 3.2.1.1) precursor - Bacillus stearothermophilus (strain DY-5) plas

N:Alternate names: 1,4-alpha-D-glucan glucanohydrolase
 C:Species: Bacillus stearothermophilus
 C:Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 16-Feb-1997
 C:Accession: A91999; B91999; A91804; A00845
 R:Iihara, H.; Sasaki, T.; Tsuboi, A.; Yamagata, H.; Tsukagoshi, N.; Ueda, S.
 J. Biochem. 98, 95-103, 1985

A:Title: Complete nucleotide sequence of a thermophilic alpha-amylase gene: homology
 A:Reference number: A91999; MUID:86008166; PMID:3876333
 A:Accession: A91999
 A:Molecule type: DNA
 A:Residues: 1-548 <IH1>
 A:Cross-references: GB:X02769

A:Experimental source: plasmid PH1300 from strain DY-5
 A:Accession: B91999
 A:Molecule type: protein
 A:Residues: 35-48 <IH2>
 A:Experimental source: strain DY-5

R:Tsukagoshi, N.; Iritani, S.; Sasaki, T.; Takemura, T.; Ihara, H.; Idota, Y.; Yamaga
 U. Bacteriol. 164, 1182-1187, 1985
 A:Title: Efficient synthesis and secretion of a thermophilic alpha-amylase by protein
 F:1-34/Domain: signal sequence #status predicted <SIG>
 F:35-548/Product: alpha-amylase #status predicted <AMT>
 F:235-368/Domain: alpha-amylase core homology <AMT>
 F:139,237,272/Binding site: calcium (Asp, Asp, His) #status predicted
 F:268,298,365/Active site: Asp, Glu, Asp #status predicted

A:Residues: 1-29, 'Q', 31-75, 'W', 77-122 <TSD>
 C:Comment: Alpha-amylase genes have been found on plasmids and in multiple copies on
 C:Genetics:
 A:Genome: plasmid
 A:Start codon: GTG
 C:Function:

A:Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds
 A:Pathway: glycogen/starch degradation
 C:Superfamily: alpha-amylase, amyloliquefaciens type; alpha-amylase core homology
 C:Keywords: extracellular protein; glycosidase; heat-stable protein; hydrolase; polys

F:1-34/Domain: signal sequence #status predicted <SIG>
 F:35-548/Product: alpha-amylase #status predicted <AMT>
 F:235-368/Domain: alpha-amylase core homology <AMT>
 F:139,237,272/Binding site: calcium (Asp, Asp, His) #status predicted
 F:268,298,365/Active site: Asp, Glu, Asp #status predicted

Query Match 6.1%; Score 105; DB 1; Length 548;
 Best Local Similarity 19.3%; Pred. No. 1.2; Indels 118; Gaps 17;
 Matches 68; Conservative 50; Mismatches 117; Indels 118; Gaps 17;

QY 50 SYKEFAESALEETRRRIYLPESHLOEKYSSMTGLDGKTELQMLAKSSKIRLLRMAIEN 109
 DB 229 NVDYLMVADLDHDEHVEVTELKNMGKWTNTNIDG-----FRIDGLKH 272
 QY 110 ETMOVF-DPAGFMEPEYDPIFCANFTSTNNIVLNLPLHQLDQD 156
 DB 273 IKFSEFPDMLSVRSQGTGKPLETVEYWSYDINKLHNYITRTDGTMSLFD-PLH----- 326
 QY 157 TDYQDKYV-----NKMSIYHKYAEF-----PMGSKL 184
 DB 327 -----NKFTASKSGGAFDMRTLMTNTLMKDDPTLAVTFVNDHDEPQALQSWDPM----- 378
 QY 185 TGESIRKFSPLVMTTRFSSSEKHKALFSAFLEYQAWLEMTIQVREMEPSHVANCEA 244
 DB 379 -----FKPLA-YAFILTRQEGPCVF--YGDYIGIPQYNIPSLKSKIDPLIARDYA 428
 QY 245 -QHKYLT-----WRAQ---KDPGHGLKRL-----VGEAKAKELLRDPLEN 281
 DB 429 YGTQHDYLDHSDIIGMTREGTEKPGSGLAALITDGAQSKMMVYKQHGAKVYFDLTGN 488
 QY 282 GVDELGRTFIDFPEYQTEDEGTVS-----DKRSIIIGKSYETRRPMDLTGQFI 328

Db

489 RSDVTYVINS--DGMGEFKNVGSVSWVPKRTVTSTIARPLITTRP--TGEFV 537

RESULT 8

A24549

alpha-amylase (EC 3.2.1.1) precursor - *Bacillus stearothermophilus* (strain NZ-3)
 N:Alternate names: 1,4-alpha-D-glucan glucanohydrolase
 C:Species: *Bacillus stearothermophilus*
 C>Date: 30-Jun-1988 #sequence_revision 18-Aug-1995 #text_change 18-Jun-1999
 C:Accession: A24549; 139501; 139770
 R:Gray, G.L.; Malnzer, S.E.; Rey, M.W.; Lama, M.H.; Kindle, K.L.; Carmona, C.; Requet, J.
 J.Bacteriol. 166, 635-643, 1988
 A:Title: Structural genes encoding the thermophilic alpha-amylases of *Bacillus stearothermophilus*
 A:Reference number: A01817; M01D:86195857; PMID:3009417
 A:Accession: A24549
 A:Molecule type: DNA
 A:Residues: 1-549 <GR>

A:Cross-references: GB:M13255; NID:g142512; PIDN:AAA2241.1; PID:g142513
 A:Experimental source: genomic DNA of strain NZ-3
 R:Satoh, H.; Nishida, H.; Isono, K.
 J.Bacteriol. 170, 1034-1040, 1988
 A:Title: Evidence for movement of the alpha-amylase gene into two phylogenetically distant
 A:Reference number: 139501; M01D:86139156; PMID:3257753
 A:Accession: 139501
 A:Molecule type: DNA
 A:Residues: 536-549 <RES>
 A:Cross-references: GB:M29577; NID:g142476; PIDN:AAA2225.1; PID:g142478
 A:Experimental source: strain DY-5
 A:Accession: 139770
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 536-549 <RE2>
 A:Cross-references: GB:M29578; NID:g142484; PIDN:AAA2228.1; PID:g142486
 A:Experimental source: strain 799
 C:Comment: Alpha-amylase genes have been found on plasmids and in multiple copies on the
 C:Genetics:
 A:Start codon: GTG
 C:Function:

A:Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds
 A:Pathway: glycoen/starch degradation
 C:Superfamily: alpha-amylase, amyloliofaciens type; alpha-amylase core homology
 C:Keywords: extracellular protein; glycosidase; heat-stable protein; hydrolase; polysac
 F:1-34/Domain: signal sequence #status predicted <SIG>
 F:235-368/Product: alpha-amylase #status predicted <SIG>
 F:139,237,272/Binding site: calcium (Asp, Asp, His) #status predicted <CAT>
 F:268,298,365/Active site: Asp, Glu, Asp #status predicted

Query Match

Best Local Similarity 6.1%; Score 104.5; DB 1; Length 549;
 Matches 69; Conservative 49; Mismatches 117; Indels 119; Gaps 17;

QY 50 SYKEFAESALETRKRYLEPSHLOEKYSMTGLDQKTELQMLAKRSKIRLRSMALEN 109
 DB 229 NYDYLAMADLMQHEVEVTLKMGKMYNTNIDQ-----FRIDAVKH 272
 QY 110 EFMQVF--DFAFMPEPEVDPRFCA-----NEFTSTNVIIVLDLNPJHQLTDQ 156
 DB 273 IKSEFPDMLSYRSQTKRPLFVGEWYSYDINKLHNITFTKNGMLFDP-PLH----- 326
 QY 157 TDYQDKRY-----NKINSYHKYETP-----PWGGR 183
 DB 327 ---NKFYASKSGAFDMSTLMNNTLMKQDPLAVTFVVDNHDTEPQALQSWVDW--- 379
 QY 184 LTGESIKFSPPLVMTFRSSSEKHKALFSAFLTYOAMLEMTIOVREMEPSHVANCE 243
 DB 380 -----FKPLA-YAFILTRQEGYPCVF--YGDYGIPOYNIPSLKSIDPLLRARY 428
 QY 244 A---OHKYLIT-----WAAQ---KDPGHGLKRLR-----VGEAKAKELLADPLF 280
 DB 429 AVGTGHDLDLSDIITGTRGVTREKPGSGIALITDGPGRKMYVGVKOHAKYVYDULG 488

QY

281 NGVDELGKTFTIDYPPETQEDGTVS-----DKRSITGSKSYTRPMDLTGOFI 328
 DB 489 NRSPTYVINS--DGMGEFKNVGSVSWVPKRTVTSTIARPLITTRP--TGEFV 538

RESULT 9

T30905

hypothetical protein - *Caldocellum saccharolyticum* (fragment)
 C:Species: *Caldocellum saccharolyticum*
 C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
 C:Accession: T30905
 R:Te'o, J.; V.S.; Gibbs, M.D.; Saul, D.J.; Bergquist, P.L.
 Submitted to the EMBL Data Library, May 1997
 A:Reference number: 220932
 A:Accession: T30905
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-311 <TR>
 A:Cross-references: EMBL:AF005383; NID:g2645411; PID:g2645412; PIDN:AA87367.1

Query Match 5.9%; Score 102.5; DB 2; Length 311;
 Best Local Similarity 19.9%; Pred. No. 0.92;
 Matches 66; Conservative 44; Mismatches 106; Indels 115; Gaps 15;

QY 9 FSIGCFKAPNPVYLISAPN-----KINFTRRRKRLKRSVAVYKFAE 56
 DB 62 YSPATCYPDQDYVPEFMLESPNYGFRVSAIPSLNT-----KNISISSIKMLYSDFCT 117
 QY 57 SALETRKRYLEPSHLOEKYSMTGLDQKTELQMLAF--KSKIRLL-----RSMALENET 111
 DB 118 RQINRLQTEKIFE-----QIYGIKNRHYNLLLEPNCGLITNAS 157
 QY 112 MQVFDFAGFMEPEVDPRFCAFPFTSTNVIIVLDLNPJHQLTDQYQDKRYKIMSTY 171
 DB 158 EIV-----DIPIYSKFAIE-----DIPIFYQVNLGY 186
 QY 172 -----HKAEFPWGRKL--TGESIKFSPPLVMTFRSSSKHKALFSAFLLEYQA 221
 DB 187 IPYTPSVNIDYSSSEMLKLVLESGSYIKF-----TWARRDELEKETICDIYSSNYKL 241
 QY 222 WLENTIOVREMP--SHRANCEQHKYLTNRQKDPGHLRLVGEAKAKELLADPL 279
 DB 242 WLDLKLKVKKEIYPLVSKK-----NKKFLERLKE-GLVLYK----- 279
 QY 280 FNGVDDELGKTFTIDYPPETQEDGTVSDKRS 310
 DB 280 FEG-----GTETIINTSTSDORVNTVYKARN 306

RESULT 10

B97866

DNA primase (EC 2.7.7.-) [imported] - *Rickettsia conorii* (strain Malish 7)
 C:Species: *Rickettsia conorii*
 C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
 C:Accession: B97866
 R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.;
 Science 293, 2093-2098, 2001
 A:Title: Mechanisms of Evolution in *Rickettsia conorii* and *Rickettsia prowazekii*.
 A:Reference number: A97700; M01D:21442074; PMID:11557893
 A:Accession: B97866
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-595 <RUP>
 A:Cross-references: GB:AE006914; PIDN:AA03868.1; PID:g15620472; GSPDB:GN00173
 C:Genetics:
 A:Gene: dnag
 C:Superfamily: DNA primase
 C:Keywords: nucleotidyltransferase

Query Match 5.5%; Score 95.5; DB 2; Length 595;
 Best Local Similarity 20.0%; Pred. No. 8.2;
 Matches 77; Conservative 60; Mismatches 147; Indels 101; Gaps 17;

QY 1 MALSMFSGISGSC-FKAPNPVLISAPNKINFTLRRKK-----RELLR 45
 DB 25 VALTRSGNYVGLCPHOKETP-----SFYVNSGRFFCYFCGCAADVIKFTSN 74
 QY 46 VSAVSKFEAEALAEETRRRIYLEPSHLOEKYSMTGLDGTQLMIAFKSSK-----IRL 101
 DB 75 ISGLYNEISAKILADYGEIRPLAKOKEFYESEELINILIELANKFFRTQTLPELHY 134
 QY 102 LRSMALNETMOVDFAGMEPEYDTPICANFTSTNNIYVLDINPLHQLTDQTDYD 161
 DB 135 LHERGTETETVEFESI-GE-APKNNK---FEKFFHKKNDIILKLGAGLIGKRENEIYN 189
 QY 162 KYNNKI-MSIYHKYAEFPFGMGKRLNGESIKFSPLMVWRFPSSSK---KHKALFSAF-- 215
 DB 190 LFSNRITPIRNIYKNGVGRVJGEBGLPKYLNSEPTTVFQKSELYGEHKAISSTYK 249
 QY 216 -----LEYYOAWLEMTIOVREMEPSHVANCEAOHKYLTWPAOKDPCHGLK 263
 DB 250 NMSILVEGFVIALHQAQSFETV-----ASLGTSTVENHHLKLMRA---GDEIIL 297
 QY 264 RLVEGE-ARKKELLRDE-----LFNG-----VDELGTRTFIDYFPEYQTEDEG 303
 DB 298 CLDGNAGIKASIRITNIALPLINSEKKISFIRLPSGLDPPDAVNKGADFFAK----- 351
 QY 304 TVSDKRSTI-----GKSYET 318
 DB 352 -LIDKRISLSEIMIMHTEYSGKSFKT 375

RESULT 11

F70177
 transcription-repair coupling factor (mfd) homolog - Lyme disease spirochete
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
 C:Accession: F70177
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Whitte-
 son, D.; Peterson, J.; Kariya, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
 ; Bowman, C.; Garland, S.; Fujil, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; MID:98065943; PMID:9403685
 A:Accession: F70177
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1125 <KLE>
 A:Cross-references: GB:AE001163; GB:AE000783; NID:g2688541; PIDN:AA066973.1; PID:g268854
 A:Experimental source: strain B31
 C:Superfamily: transcription-repair coupling protein
 C:Keywords: ATP; DNA repair; leucine zipper; nucleotide binding; P-loop; transcription
 F:610-617/Region: nucleotide-binding motif A (P-loop)
 F:707-714/Region: nucleotide-binding motif B
 F:711-714/Region: DEAD/H motif #status atypical

Query Match 5.5%; Score 95.5; DB 1; Length 1125;

Best local similarity 22.0%; Pred. No. 19; Mismatches 107; Indels 45; Gaps 11;

Matches 55; Conservative 43; Mismatches 107; Indels 45; Gaps 11;
 QY 64 KRIVLEPSHLOEKYSMTGLDGTQLMIAF-KSKIRILRSMALNETMOVDFAGME 122
 DB 133 KMTLLKNYKIKKNNININADIEKTLITLGEYKTLAVTIPGEFTYVAGELIIDIPFG----- 188
 QY 123 PEYDTPICANFTSTNNIYVLDINPLHQLTDQD-----YDKYV----- 165
 DB 189 --EONPIRIALNFKIE--EIKKFNPLQLKHNDEILFEQLPKKEIIMDKTITMLTKT 243
 QY 166 KIMSI-YHKYAEFPFGMGKRLNGESIKFSPLMVWRFPSSSKRKLFSALFXYQAMLE 224
 DB 244 KIKSVYKILLELDF--KKEKTEEMFPLVANTYLGDEIEKHPITVNEFINNEKEIE 301
 QY 225 MTIOVREMEPSHVANCEAOHKYLTWPAOKDPCHGLKRLVGEAKAKELLDFLNGVD 284

DB 302 ---KIHQYEKLYKEAE-EAGKNII-----DP-----KRILLNTYFNLKSDVLESKIK 346
 QY 285 ELGTRTFIDY 294
 DB 347 SLKSKETIEF 356

RESULT 12

B35149
 IpaH protein - Shigella flexneri
 C:Species: Shigella flexneri
 C:Date: 03-Aug-1990 #sequence_revision 12-Apr-1996 #text_change 03-Dec-1999
 C:Accession: S18248; B35149
 R:Venkatesan, M.M.; Buysse, J.M.; Hartman, A.B.
 Mol. Microbiol. 5, 2435-2445, 1991
 A:Title: Sequence variation in two IpaH genes of Shigella flexneri 5 and homology to
 A:Reference number: S18248; MID:92167809; PMID:1791758
 A:Accession: S18248
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-574 <VEN>
 A:Cross-references: GB:M76445; NID:g152751; PIDN:AAA26528.1; PID:g152752
 R:Hartman, A.B.; Venkatesan, M.; Oaks, E.V.; Buysse, J.M.
 J. Bacteriol. 172, 1905-1915, 1990
 A:Title: Sequence and molecular characterization of a multicopy invasion plasmid anti-
 A:Reference number: A35149; MID:90202708; PMID:1690703
 A:Accession: B35149
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-208 <HAR>
 A:Cross-references: GB:M32063; NID:g152746; PIDN:AAA26527.1; PID:g152748
 C:Genetics:
 A:Gene: IpaH

Query Match 5.5%; Score 95; DB 2; Length 574;

Best local similarity 21.4%; Pred. No. 8.5; Mismatches 85; Conservative 41; Mismatches 145; Indels 126; Gaps 18;

QY 3 LSNREFSISGCKRANPVL--ISASPKINTLRRKKRRLRLRSANSYKFAFSALE 60
 DB 127 LNIISFNRL-SCUPS-LPPYLOSLSARFNSLETLPSTLTTLRI-----E 171
 QY 61 ETRKRIVLE-PSHLOEKYSMTGLDGTKE---LQMLAFKSKIRILRSMALNETMOYF 115
 DB 172 GNRLLTVLPRLPHLOGLFVSGNLOGLPEPPOSGLKLVKGENDLRKLSRLOGLALDVS 231
 QY 116 DFAGMEPEYDTPICANFTSTNNIYVLDL-----NP-----LHQLTDQTDY 159
 DB 232 N-----NLLTSLPENITLPICTNWNISGNPLSTRVLSGLRLTSSPDY 275
 QY 160 QDKYVYKIMSIYHKYAEFPFGMGKRLNGESIKFSPV-----MNTRES 202
 DB 276 HGP-----QIYFSMD-----GQONTLRLPLADAVTANFPENKOSDVSIWHA- 319
 QY 203 SSKKHALFSAFL-----EYQAWLEMTIOVREMEPSHVRA-----N 241
 DB 320 -EHEEIANFTSAFLDLSDTVSARNTSGREFQVAAMLERLSAELRQGSFAVAADATES 378
 QY 242 CEAOHKYLTWPAOK-----DPGHGLKRLVGEAKAKELLDFLNGVDDELGT 288
 DB 379 CE-DRAVALMNNLRKTLVHQASEGLFENDDTGALLSLGRMFRLEILEDI---ARDKVR 434
 QY 289 KTFIDYFPEYQTEDEGVSDKRSITGKSYETRPMDLNG 325
 DB 435 LHFVDEIEYVLAETQMLAEKQLSTAVKEMRFYGVSG 471

RESULT 13

T33741
 DNA-binding protein Iag-1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T33741

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 17:19:10 ; Search time 21 Seconds

(without alignments)
649.796 Million cell updates/sec

Title: US-09-870-406A-33

Perfect score: 1724
Sequence: 1 MALSMERGFSGSCFRAPNP.....SIIGSYETRPMDLTGQFIG 329

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	210.5	12.2	YCP3_SYNPY	Q02190 synechococ
2	158	9.2	YCP2_SYNPY	Q02189 synechococ
3	95.5	5.5	MEP_BORBU	O51568 borrelia bu
4	95	5.5	IPPA4_SHIFL	P18009 shigella fl
5	95	5.5	IGGL_YEAST	P36143 saccharomyc
6	93.5	5.4	VD36_AQUAE	O67355 aquifex aeo
7	92	5.3	CISTY_HUMAN	O75390 homo sapien
8	90.5	5.2	SBQC_LACIA	Q95fz0 lactococcus
9	90.5	5.2	FERI_CAEBL	Q17388 caenorhabdi
10	90	5.2	YG4B_YEAST	P46951 saccharomyc
11	89	5.2	DPO3_MYCSE	P47277 mycoplasma
12	88.5	5.1	APB_CHICK	P11682 gallus gall
13	88.5	5.1	AMV_BACST	P06279 bacillus st
14	88.5	5.1	DEP3_YEAST	Q08225 saccharomyc
15	88.5	5.1	ODOL_RICPR	Q14690 homo sapien
16	88.5	5.1	RRP5_HUMAN	P57485 buchnera ap
17	88	5.1	GSN1_BUCAL	P10959 rattus norv
18	88	5.1	EST1_RAT	Q99qb3 mycoplasma
19	87.5	5.1	PRIM_MYCPU	P45181 haemophilus
20	87	5.0	POOL_HAEIN	Q95986 chlamydia p
21	87	5.0	PKN1_CHLPN	O03701 homo sapien
22	86	5.0	CBF_HUMAN	P57316 buchnera ap
23	85.5	5.0	MORE_HUMAN	P71561 mycobacteri
24	85.5	5.0	771	O51578 borrelia bu
25	85.5	5.0	EXSB_BORBU	Q14765 homo sapien
26	85	4.9	STRA4_HUMAN	O03660 saccharomyc
27	85	4.9	TR13_YEAST	P00889 sus scrofa
28	84	4.9	CISY_PIG	O12663 saccharomyc
29	84	4.9	GIN4_YEAST	P14268 b graminicid
30	84	4.9	GRSB_BACDR	P81434 buchnera ap
31	83.5	4.8	ST5_BUCAP	P33538 neurospora
32	83	4.8	DPOV_NEUDIN	Q12451 saccharomyc
33	83	4.8	OSH2_YEAST	

34	82.5	4.8	506	1	SYK_XYLA	O9p66 xyloella fas
35	82	4.8	291	1	ERA_CAMIE	O9ph11 campylobact
36	82	4.8	558	1	CX1B_PARE	P96802 paracoccus
37	82	4.8	565	1	HOFU_PSEX	O9agu4 pseudomonas
38	82	4.8	738	1	ECT2_MOUSE	O07139 mus musculu
39	82	4.8	1835	1	DUR1_YEAST	P32528 saccharomyc
40	81.5	4.7	203	1	YPT1_NEUR	P33723 neurospora
41	81.5	4.7	207	1	ALAH_MUSCR	P21352 mus caroli
42	81.5	4.7	370	1	ORC_ASPNG	P11066 aspergillus
43	81.5	4.7	430	1	STD_ARCFU	O29342 archaeoglob
44	81.5	4.7	491	1	C2F2_MOUSE	P33267 mus musculu
45	81.5	4.7	932	1	SSN1_PASHA	P31631 pasteurella

ALIGNMENTS

RESULT 1	ID	YCP3_SYNPY	STANDARD:	PRT:	257 AA.
AC	002190	1993 (Rel. 26, Created)			
DT	01-JUL-1993	(Rel. 26, Last sequence update)			
DT	01-JUL-1993	(Rel. 26, Last sequence update)			
DT	01-OCT-1993	(Rel. 27, Last annotation update)			
DE	Hypothetical 29.3 kDa protein in cpeb 3' region (ORF257).				
OS	Synechococcus sp. (strain WH8020).				
OC	Bacteria: Cyanobacteria; Chroococcales; Synechococcus.				
OX	NCBI_TaxID=32052;				
RM	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=93144698; PubMed=8425055;				
RA	de Lorimer R., Wilbanks S.M., Glazer A.N.;				
RT	"Genes of the R-phycocyanin II locus of marine Synechococcus spp., and comparison of protein-chromophore interactions in phycocyanins differing in bilin composition.";				
RL	Plant Mol. Biol. 21:225-237(1993).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=93123238; PubMed=8419325;				
RA	Wilbanks S.M., Glazer A.N.;				
RT	"Rod structure of a phycoerythrin II-containing phycobillosome. I. Organization and sequence of the gene cluster encoding the major phycobilliprotein rod components in the genome of marine Synechococcus sp. WH8020.";				
RL	J. Biol. Chem. 268:1226-1241(1993).				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	EMBL: M95288; AAA27344.1; -				
DR	PIR: B46448; B46448.				
DR	PIR: S31059; S31059.				
KW	Hypothetical protein; Phycobillosome.				
SQ	SEQUENCE 257 AA; 29265 MW; AE5926A1B4388393 CRC64;				
Query Match	12.2%; Score 210.5; DB 1; Length 257;				
Best local similarity	24.3%; Pred. No. 8.7e-10;				
Matches	63; Conservative 55; Mismatches 126; Indels 15; Gaps 8;				
OY	34 TLRRRRKRLRVSASVYKEPAESALEETRRKRIYLEPSHLDKY---SSMGLDGKT--- 87				
DB	2 TNORRKSYPVNIIEGSMQPLEDAIKRL-EGIANEPYVPDRFQORDQGSKSKSIPV 60				
OY	88 ELQMAFASSKTRILRLSMAIE-NEIMOVFDPAFGHEPYPDPICANFTSTNNVIYID 146				
DB	61 TTATWACKTERFROVRAACVSAGSAAVSLNIPVINDKSYGLPEFGDVLTPPAGLLAMD 120				
OY	147 LNPILHQLTDQNDYODKRYNKMISYHKAETFPWCGKITGESIKFFSPDLVMTTRSSSKE 206				

```

Db      121 LOPARK-DEV-HTTPWDRILPIFERMRDLPYGGPIPEBAOPFSPGFLMTRLPJEE 178
OY      207 KHALES-ATLEYOAWLEMTIQVREMEPSVRANCAQKHLYLWRKQDPGHL 262
Db      179 GDELQSLVRAFNNDYLDLYELASA-ERVTERSEVLLQGRKYTDYRAEKDPARCM 237
OY      263 KRLVGEAKKELNDFLN 281
Db      238 TRFGSEWTEAVYHTVLED 256

RESULT 2
YCP2_SYNPY STANDARD; PRT; 236 AA.
ID YCP2_SYNPY
AC 002189;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE 01-JUL-1993 (Rel. 26, Last annotation update)
DE Hypothetical 27.2 KDa protein in cpes 3 region (ORF236).
OS Synechococcus sp. (strain WH8020).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
CX NCBI_TaxID=32052;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE-93144696; PubMed-8425055;
RA de Lotimier R., Wilbanks S.M., Glazer A.N.;
RT "Genes of the R-phycoerythrin II locus of marine Synechococcus spp.,
RT and comparison of protein-chromophore interactions in phycoerythrin
RT differing in bilin composition."
RL Plant Mol. Biol. 21:225-237(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; M95288; AAA27343.1; -
DR PIR; S31058; S31058.
KW Hypothetical protein; Phycobilisome.
SQ SEQUENCE 236 AA; 27244 MW; 27CCEB17B50BD8 CRC64;

Query Match 9.28; Score 158; DB 1; Length 236;
Best Local Similarity 24.48; Pred. No. 1.2e-05;
Matches 60; Conservative 41; Mismatches 89; Indels 56; Gaps 11;

OY 71 SHLOERYSSMT-----GLDKTELQMLAFSSK-----RLRSM 105
Db 4 SFINELHSIDITKGGSPLEPLE-----ECRSSKSSSVYQSLMDVPGFRMRVYRLD 58
OY 106 ALENEMTOVDFAGFMEPEYDFIPFCAN--FRTSTVNVIVLNLPLHOLTDQTDQDKY 163
Db 59 A--GDSLOVNSVAVYDYNVDHPLMGVDLIMFARQKIVAVLDPOVLQ---DKYLDRI 113
OY 164 YKIMSLYKRYATFTFPGCKLIG-ESIKFESP--LVMTFRSSSKKHKAL-----FSA 214
Db 114 FSGIKELNORFFPD-----LNGEETMRSEFDPQYFSSWMLFCRGAEQADSLPKAFSA 166
OY 215 FLEYYOAWLEMTIQVREMEPSVRANCAQKHLYLWRKQDPGHLKRLVGEAKAKEL 274
Db 167 FLKAYVDLHDMNAKSTIPPEEVK--NLQDKYDYSARDPANGLFTSHGKDWSNRF 223
OY 275 LRDFLP 280
Db 224 LRDFLP 229

RESULT 3
MFD_BORBU STANDARD; PRT; 1125 AA.
ID MFD_BORBU

```

```

AC 051568;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription-repair coupling factor (TRCF).
DE MFD OR BB0623.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
CX NCBI_TaxID=139;
RN 11
RP SEQUENCE FROM N.A.
RX STRAIN-ATCC 35210 / B31;
RX MEDLINE-98065943; PubMed-9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Keislavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman C.,
RA Uterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi."
RL Nature 390:580-586(1997).
CC CC
CC -1- FUNCTION: NECESSARY FOR STRAND-SPECIFIC REPAIR. A LESION IN THE
CC TEMPLATE STRAND BLOCKS THE RNA POLYMERASE COMPLEX (RNAP). THE
CC RNAP-DNA-RNA COMPLEX IS SPECIFICALLY RECOGNIZED BY TRCF WHICH
CC RELEASES RNAP AND THE TRUNCATED TRANSCRIPT; THE TRCF MAY REPLACE
CC RNAP AT THE LESION SITE AND THEN RECRUIT THE UVRA/B/C REPAIR
CC SYSTEM (BY SIMILARITY).
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE UVRA FAMILY.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE HELICASE
CC FAMILY. REG SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; AE001163; AAC66973.1; -
DR TIGR; BB0623; -
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR004576; Mfd.
DR InterPro: IPR005118; TRCF.
DR Pfam: PF00270; DEAD; 1.
DR Pfam: PF03461; TRCF; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
DR TIGR: TIGR00580; mfd; 1.
KW Helicase; DNA repair; ATP-binding; Complete proteome.
FT NP_BIND 610 617
FT SITE 711 714 DEO BOX.
SQ SEQUENCE 1125 AA; 130729 MW; 116PFB0E9024539 CRC64;

Query Match 5.58; Score 95.5; DB 1; Length 1125;
Best Local Similarity 22.08; Pred. No. 8.4;
Matches 55; Conservative 43; Mismatches 107; Indels 45; Gaps 11;

OY 64 KRIVLEPSHLOERYSSMTGLDKTELQMLAF-KSSKIRLRLSMALENEMTOVDFDAGWE 122
Db 133 KNTILKNIYIKKNTNININADIEKTLITLIGYEKTLKVTTPGEEYVKGELIDIDYPPG 188
OY 123 PEYDFIPFCANFTSTVNVIVLNLPLHOLTDQTD-----YDKYYN----- 165
Db 189 --EQNPRIALNFDEI--EIKKNPPLQKHDNEIIEFQILPKKELIMDKTINTILKT 243
OY 166 KIMSL-YKRYATFTFPGCKLIGESIKFESP--LVMTFRSSSKKHKALFSAFLEYOAWLE 224

```

DB 244 KIKSVYKKILEELDF--KKEETKEEMFYPLVANTYLGDEIEKHTPIVNFENNEKEIE 301
 OY 225 MTIOVREEMPSHRANCEAOHKYLTWRAQKDPCHGLKLVGEAKAKELRLDPLFNGVD 284
 DB 302 ---KIHQEVYKLYKEAE-EAGKNII-----DP-----KRLINKTKTNLSKDVLEFSKIK 346
 OY 285 ELGKTFIDY 294
 DB 347 SLKSKETIEF 356

RESULT 4
 ID IPA4_SHIFL STANDARD: PRT: 574 AA.
 AC P18009:
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE 65.4 kDa antigen.
 GN IPAH 4.5.
 OS Shigella flexneri.
 OC Plasmid 210 kb invasion pWR100.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Shigella.
 OC NCBI_TaxID=623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=M90T / Serotype 5;
 RX MEDLINE=92167809; PubMed=1791758;
 RA Venkatesan M.M., Buysse J.M., Hartman A.B.;
 RT "Sequence variation in two ipah genes of Shigella flexneri 5 and
 RL homology to the LRG-like family of proteins.";
 RN Mol. Microbiol. 5:2435-2445(1991).
 RN [2]
 RP SEQUENCE OF 1-208 FROM N.A.
 RC STRAIN=M90T / Serotype 5;
 RX MEDLINE=90202708; PubMed=1690703;
 RA Hartman A.B., Venkatesan M.M., Oaks E.V., Buysse J.M.;
 RT "Sequence and molecular characterization of a multicopy invasion
 RL plasmid antigen gene, ipah, of Shigella flexneri.";
 RN J. Bacteriol. 172:1905-1915(1990).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M76445; AAA26528.1; -;
 DR EMBL: M32063; AAA26527.1; -;
 DR EMBL: B35149; B35149.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR003592; LRR_out.
 DR Pfam: PF00560; LRR: 4.
 DR SMART: SMO0370; LRR: 8.
 KW Antigen; Plasmid; Virulence; Repeat; Multigene family.
 FT DOMAIN 74 228
 FT 8 X 14 AA APPROXIMATE TANDEM REPEATS OF
 FT L-X(2)-L-P-X-L-P-X(2)-L-X(2)-L.
 FT REPEAT 74 87
 FT REPEAT 94 107
 FT REPEAT 114 127
 FT REPEAT 135 148
 FT REPEAT 155 168
 FT REPEAT 175 188
 FT REPEAT 195 208
 FT REPEAT 215 228
 SQ SEQUENCE 574 AA: 65408 MM: 01E083B3446C97B2 CRC64;
 Query Match 5.58; Score 95; DB 1; Length 574;
 Best Local Similarity 21.48; Pred. No. 3.9;
 Matches 85; Conservative 41; Mismatches 145; Indels 126; Gaps 18;

OY 3 LSMERFGSIGSCFAPNPVVL--ISAPNKNIFLRRKKRFLRLSAVSAYKEFAESALE 60
 DB 127 LNISFNRL-SCLPS-LPPYLOSARNSLETLEPDELSTLTILTI-----E 171
 OY 61 ETRKRIVLE-PSHLQEKYSMTGLDGKE---LQMLAFKSKIRLRSMAIENETMOVF 115
 DB 172 GNRLLVLEPLRHLRQELFVSGNRQLELPEFPQSLKYLKVGENGRLRLPQELLALDVS 231
 OY 116 DPAGFMEPEYDTPICANFTSTNNIVYLDI-----NP-----LHQLTDQDY 159
 DB 232 N-----NLTSLEPENITLTPTCTNWNISGNPLSTRLVLSQLRLTSSPDY 275
 OY 160 QDKYNNKIMSTYHKAEFFPPMGKLTGSEINFSLY-----WMTFS 202
 DB 276 HGP-----QIFYSKSD-----GOQNTLHRLADAVYAMPENKOSDYSQIWHAF 319
 OY 203 SSKERKALFSNFL-----EYQAMLEMTIOVREEMPSHYRA-----N 241
 DB 320 -EHEHNTYFSAFDLRLSDTVSARTSGFREGVAMLEKLSASALRQOSFAVADATES 378
 OY 242 CEAOHKYLTWRAQK-----DPGHGLKRLVGEAKAKELRLDPLFNGVDELGT 288
 DB 379 CE-DRVALTWNNLRKTLTVHQASEGFLFNDTGALLSLGREMFRLEILEDI--ARDKYRT 434
 OY 289 KTFIDYPEYQTEDEGTVDKRSIIIGKSYETRPMDLTG 325
 DB 435 LHFVDEIVYLAFOTMLAEKQLSTAVAKEMRYGVSG 471

RESULT 5
 ID GLG1_YEAST STANDARD: PRT: 618 AA.
 AC P36143:
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Glycogen synthesis initiator protein GLG1.
 GN GLG1 OR YKR058W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
 OC NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=96069371; PubMed=8524228;
 RX Cheng C., Mu J., Faras I., Huang D., Goebel M.G., Roach P.J.;
 RT "Requirement of the self-glucosylating initiator proteins Glg1p and
 RL Glg2p for glycogen accumulation in Saccharomyces cerevisiae.";
 RN Mol. Cell. Biol. 15:6632-6640(1995).
 RN [2]
 RP SEQUENCE OF 139-480 FROM N.A.
 RC STRAIN=S288C;
 RA van Vliet-Reedijk J.C., Planta R.J.;
 RL Submitted (MAR-1994) to the EMBL/Genbank/DBJ databases.
 CC -!- FUNCTION: SELF-GLUCOSYLATING INITIATOR OF GLYCOGEN SYNTHESIS. ACT
 CC AS A PRIMER FOR THE ELONGATION REACTION CATALYZED BY GLYCOGEN
 CC SYNTHASE.
 CC -!- SIMILARITY: BELONGS TO THE GLYCOGENIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U25546; AA91646.1; -;
 DR EMBL: Z28283; CAAB2136.1; -;
 DR PIR: S38134; S38134.
 DR SCD: S0001766; GLG1.
 DR InterPro: IPR002495; GL_8.

DR Pfam; PF01501; Glyco_transf_8; 1.
 KW Glycogen biosynthesis.
 SO SEQUENCE 618 AA; 69912 MW; E5422F6ACED9DCA CRC64;

Query Match 5.5%; Score 95; DB 1; Length 618;
 Best Local Similarity 23.5%; Pred. No. 4.3;

Matches 54; Conservative 37; Mismatches 85; Indels 54; Gaps 12;

QY 53 EFASALEEPRKRIVL-EPHIOE---KYSSMTGLDQKTELQMAFKSSKIRLLR--- 103
 DB 59 ELAKNILOSITKTLVLEPLNCQESIQKNSENALLERPELSALLKARIMELTQFEQV 118
 QY 104 -----SMAIENEMQVDF-----AGFME--PEYTPIFCAN 133
 DB 119 LYIDSDLPKKEFLKLEPDKMSQTSQVGAIDIGWPMENSGVMMLIPDADASYLQN 178
 QY 134 FFTSTNVIYVLDNPHHQLTDQ---TD--YQDKYKINSIYHKAETFPWGKLGES 188
 DB 179 YIFE-NISIDSGDGIINQPNONCCTDELYKDSFSREWOLSTFTVNTIPLNGYSSPA 237
 QY 189 IKFSPPLVMTRESSSEKHK--ALFSA--FL--EYQAMLEMTIQVREE 232
 DB 238 MNYKPSIKLIHFIC---KHKPMSLMSOKNKIKNEYHDQNNVEYEEKKE 284

RESULT 6

YD36_AOUAE STANDARD; PRT; 374 AA.
 AC 067355;
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein AQ_1336.
 GN AQ_1336.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
 CC Aquifex.
 NC NCBL_TaxID=63363;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VF5;
 RX MEDLINE=9819666; PubMed=9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Snaad M.A., Keller M., Auyaj M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 aeolicus.";
 RL Nature 392:353-358(1998).
 CC -1- SIMILARITY: BELONGS TO THE UPF0103 FAMILY. AQ_1336 SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC EMBL; AE000735; AAC07322.1;
 DR InterPro; IPR002737; DUF52.
 DR Pfam; PF01875; UPF0103; 1.
 DR ProDom; PD006364; DUF52; 1.
 KW Hypothetical protein, Complete proteome.
 SO SEQUENCE 374 AA; 43275 MW; DF4966E2F756B54 CRC64;

Query Match 5.4%; Score 93.5; DB 1; Length 374;
 Best Local Similarity 22.2%; Pred. No. 3;

Matches 66; Conservative 47; Mismatches 117; Indels 67; Gaps 14;

QY 30 KINFTLRKKRRLRVSAVSYKEFASALEETRRKRIV---LEPSHQEKYSMTGLDG 85
 DB 67 EINOVINFIDENYIL-----YNERPLNKKERERIKLKGVREPSHAQVAPE-NPOL 119

QY 86 KTELQMAFKSSKIRILRSMAIENEMQVDF-----FAGMEPEYDPIF--CANFETS 137
 DB 120 KNFLEEVNKNSEKFKRGLIVPMQDLRVASGYGYSKAKENEYITVLLGSHFHE 179
 QY 138 TNVNIYVLDL-NPLHQLT-----DQTDYDKRYNKMISYHKYAE 177
 DB 180 TPFSVLPDLDRTPGLGDKVDIERVEELQKFDYDLSHDVLAQKNSHIERQITFLKY--L 237
 QY 178 FPMGKLTGSIKIFSPPLVMTRESSSEK--KHKALFSAFLLEYQAMLEMTIQVREMEPS 236
 DB 238 FP-----EVKIPAIIVSYGDKLSLKEIAHK--ITKVLDSQNPV-----TISVDFS 282
 QY 237 HVANCEAOKHYLWRAQKDPGHGLKRLVGEAKAKELRDLFNGVDELGTFTFD 293
 DB 283 HVGKRFQDPHSY-----DPSPPREYINLALAKNEE-----APNLQSDNNTRID 329

RESULT 7

CISY_HUMAN STANDARD; PRT; 466 AA.
 ID CISY_HUMAN
 AC 075390;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Citrate synthase, mitochondrial precursor (EC 4.1.3.7).
 GN CS.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NC NCBL_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=99026970; PubMed=9809442;
 RA Goldenthal M.J., Marin-Garcia J., Ananthakrishnan R.;
 RT "Cloning and molecular analysis of the human citrate synthase gene.";
 RL Genome 41:733-738(1998).
 CC -1- CATALYTIC ACTIVITY: Citrate + CoA = acetyl-CoA + H(2)O +
 CC oxaloacetate.

CC -1- PATHWAY: Tricarboxylic acid cycle.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -1- MISCELLANEOUS: CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS
 CC CAPABLE OF OXIDATIVE METABOLISM.
 CC -1- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC EMBL; AF047042; AAC25560.1;
 DR HSSP; P00889; ZCTS.
 DR GeneW; HGNC:2422; CS.
 DR MIM; 118950;
 DR InterPro; IPR002020; Citrate_synt.
 DR Pfam; PF00145; CITRINSYNTHASE.
 DR PRINTS; PR00145; CITRINSYNTHASE.
 DR PROSITE; PS00480; CITRATE_SYNTHASE; 1.
 KW Lyase; Tricarboxylic acid cycle; Mitochondrion; Transl. peptide.
 FT TRANSIT 1 27 MITOCHONDRION.
 FT CHAIN 1 27 CITRATE SYNTHASE.
 FT FT 28 466 BY SIMILARITY.
 FT ACT_SITE 301 301 BY SIMILARITY.
 FT ACT_SITE 347 347 BY SIMILARITY.
 FT ACT_SITE 402 402 BY SIMILARITY.
 SO SEQUENCE 466 AA; 51706 MW; 878841484BD7CC80 CRC64;

Query Match 5.3%; Score 92; DB 1; Length 466;
 Best Local Similarity 22.0%; Pred. No. 5.2;
 Matches 67; Conservative 38; Mismatches 104; Indels 96; Gaps 16;

```

0Y      8 GFSGSGCFK-----APNPVYLASPMKINFTLLRRKKRRLLRLVSAVSATYEFESA 58
      |||||
      95 GFSPFCQQLLPRKANGGEPIPEGLF-----WLLTGCCTPEBEYVSWL 137
      :|||
0Y      59 LEETRRIIVLEPSHLOEKYSMTGLD-KTELQMLAFSSKIRLLRSMATIENTEMOVEFD 117
      :|||
      138 SKMAKRAL-FSHV-----VTMLDNPPTLHPQSLSAAVTLNLS--ESNPAQAV-A 186
      :|||
0Y      118 AGFMEDEY-----DTPFCAN-----FFTSTNNIVLDPNLPHQLTQDTDY 159
      :|||
      187 RGSRRKYWEILEDSDVDLAKLPVAAKIYRLNLWEGSGIGALDSNLDSHNTNMLGY 246
      :|||
0Y      160 QDKRYAKINSIY--HKYAEFPWGSGKLTGESIKFFSPLVYMTFRSSSEKHKRALFSATL 216
      :|||
      247 TDHQFELMRVLYTHSDHE---GGNVSAHT---SHLY-----GSLSDPYL 287
      :|||
0Y      217 EYQALAEWITIQVREMEPSHVRAACEAQHKLYTLRAQKDPGHGLRLRYGEAKAKELL 276
      :|||
      288 SFFAAANGLA-----GPLHGLANQEV---LVWLQ-----LQKEVGKDVSDSEKLR 329
      :|||
0Y      277 DFFEN 281
      :|||
      330 DYTWN 334
      :|||

```

RESULT 8	SBCC_LACIA	STANDARD:	PRT: 1046 AA.
ID	SBCC_LACIA		
AC	Q9CFZ0;		
DT	15-JUN-2002 (Rel. 41, Created)		
DT	15-JUN-2002 (Rel. 41, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Nuclease sbccd subunit C.		
GN	SBCC OR L1321.		
OS	Lactococcus lactis (subsp. lactis) (Streptococcus lactis).		
OC	Bacteria; Filumicutes; Lactobacillales; Streptococcaceae; Lactococcus.		
OX	NCBI_TaxId:1360;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-211235183;		
RX	MEDLINE-21235186; PubMed-11337471;		
RA	Bolotin A., Wincker P., Manger S., Jallion O., Malarne K.,		
RA	Weissenbach J., Ehrlich S.D., Sorokin A.;		
RT	"The complete genome sequence of the lactic acid bacterium Lactococcus		
RT	lactis ssp. lactis IL1403. "		
RL	Genome Res. 11:731-753(2001).		
CC	-1- FUNCTION: sbccd cleaves DNA hairpin structures. These structures		
CC	can inhibit DNA replication and are intermediates in certain DNA		
CC	recombination reactions. The complex acts as a 3'->5' double		
CC	strand exonuclease that can open hairpins. It also has a 5'		
CC	-1- single-strand endonuclease activity (By similarity).		
CC	-1- SUBUNIT: Heterodimer of sbcc and sbcd (By similarity).		
CC	-1- SIMILARITY: BELONGS TO THE SMC FAMILY. SBCC SUBFAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on ways		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL: AE006364; AAK05419.1; -		
DR	InterPro: IPR003439; ABC_Transportr.		
KW	Hydrolase; Nuclease; Exonuclease; Endonuclease; DNA replication;		
KW	DNA recombination; ATP-binding; Coiled coil; Complete proteome.		
FT	NP_BIND 34	ATP (POTENTIAL).	
FT	DOMAIN 223	COILED COIL (POTENTIAL).	
FT	DOMAIN 268	COILED COIL (POTENTIAL).	
FT	DOMAIN 468	COILED COIL (POTENTIAL).	
FT	DOMAIN 543	COILED COIL (POTENTIAL).	
FT	SEQUENCE 1046 AA; 120199 MW; 8f0D0D0AC28F691 CRG64;		

Query Match	5.2%	Score 90.5:	DB 1:	Length 1046:
Best Local Similarity	19.3%	Pred. No. 19:		
Matches	67:	Conservative	66:	Matches 133: Indels 81: Gaps 14:
QY	24	ISASPNNKINFLRRKKRFFLLRVASAVSKKEFASALAEETRRKRVLPSPSLDKEYS----	78	
DB	553	IDESQKKOTVLLERKKKQLOQLTELKTSIDLKAIADFEKELSLIYSFINDYQIQIPD	612	
QY	79	-----SMTGLDGKTELQMLAEKSSKIRL--LRMAIE--NETMOYFDPAFMEPEYD	126	
DB	613	SFDEYSIDESLINTLTKSLEEBEVKNDETRKVLADLESKKLELOEKVDBEYS---NQEN	669	
QY	127	TPRICANF-----FTSTNNIVIVLDNPLQIUDQYQDKYKIKHSIYKVAETPFW	180	
DB	670	ROIENLNAETIEIGITETSYNLI-----RRKNRLMKALDFEHLSELNA-----	714	
QY	181	GGKLTGESIKFEFS-----PLVMTRFSSSKERKKALFS--AFLEYVQ-----A	221	
DB	715	--QISDIKIKISSQASLNSFESQBATLLERSANKEKIKERFSQBAFTPEFQILKEWA	772	
QY	222	WLEMTIOYREME--PSHVRCNCEAQHKYLLTWRQKDPGGLRLRLVGEAKAKELRDE	278	
DB	773	YDDDLIQISQVEVOYKADKARLKEIKNNIQOLIQNKKRPNALLI-----EEBKQOTNENY	827	
QY	279	LF-----NGVDELGKTEFIDYFPEYQREDDGVNSKRSLIGSY	316	
DB	828	VFLOKRLVASENEYEQ--AKSLTSLKRLYKQODKASKSLATIKLY	872	

```

RESULT 9
FERL_CAEEL
ID FERL_CAEEL STANDARD; PRT; 2034 AA.
AC Q17388;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sperm vesicle fusion protein ferl-1.
GN FER-1 OR F43G9.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderae; Caenorhabditia.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97318766; PubMed=9175703;
RA Achanzar W.E., Ward S.;
RT "A nematode gene required for sperm vesicle fusion."
RL J. Cell Sci. 110:1073-1081(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Bristol NZ;
RA Kershaw J.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Required for the fusion of the membranous organelles
      (MOS) with the plasma membrane, a process essential in
      spermiogenesis.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (By similarity).
CC -1- TISSUE SPECIFICITY: Exclusively expressed in the testis.
CC -1- SIMILARITY: BELONGS TO THE FERLIN FAMILY.
CC -1- SIMILARITY: CONTAINS 2 C2 DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U57652; AAB02243.1; -
DR EMBL; Z79755; CAB02109.1; -
DR WormPep; F43G9.6; CE10364.

```

Query Match	5.2%;	Score 90.5;	DB 1;	Length 2034;
Best Local Similarity	19.7%;	Pred. NO. 45;		
Matches 61;	Conservative 59;	Mismatches 114;	Indels 75;	Gaps 13;

RESULT 10	YG4B_YEAST	YG4B_YEAST	STANDARD;	PRT;	817 AA.
ID	YG4B_YEAST	STANDARD;	PRT;	817 AA.	
AC	P46951;				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	01-OCT-1996 (Rel. 34, Last annotation update)				
DE	Hypothetical 95.4 kDa protein in SNG1-PMT6 intergenic region.				
GN	YKR198W OR G7594.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycos.				
NCBI_TaxID=4932;					
OX					

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

Query Match . 5.2%; Score 90; DB 1; Length 817;
Best Local Similarity 20.0%; Pred. No. 15;
Matches 73; Conservative 59; Mismatches 129; Indels 104; Gaps 17;

RESULT 11	STANDARD:	PRT: 1451 AA.
DPO3_MYCGE		
ID DPO3_MYCGE		
AC P47277;		
DT 01-FEB-1996 (Rel. 33, Created)		
DT 01-FEB-1996 (Rel. 33, Last sequence update)		
DT 15-JUN-2002 (Rel. 41, Last annotation update)		
DE DNA polymerase III polC-type (EC 2.7.7.7) (PolIII).		
GN POLC OR MG031.		
OS Mycoplasma genitalium.		
OC Bacteria; Firmicutes; Mollitutes; Mycoplasmataceae; Mycoplasma		
NCBI_taxid=2097;		

SEQUENCE FROM N.A.
STRAIN-ATCC 33530 / G-37;
MEDLINE=96023646; PubMed=7566993;
Fraser C.M., Gockyave J.D., White O.,
Fleischmann R.D., Bult C.J., Kereyave A.R., Sutton G.,
Fritschmann J.L., Weidman J.F., Small K.V., Sandusky M.,
Nguyen D.T., Utterbaugh T.R., Saudek D.M., Phillips C.A.,
Tomb J.-F., Dougherty B.A., Bortk K.F., Hu P.-C., Lucier
Peterson S.N., Smith H.O., Hutchison C.A. III, Venter
The minimal gene complement of *Mycoplasma genitalium*.
Science 270:397-403(1995).
[2]
SEQUENCE OF 492-567, 975-1103 AND 1111-1225 FROM N.A.
STRAIN-ATCC 33530 / G-37; PubMed=8253660;
MEDLINE=94075330; PubMed=8253660;
Peterson S.N., Hu P.-C., Bortk K.F., Hutchison C.A. III, Venter

RT "A survey of the Mycoplasma genitalium genome by using random
 sequencing."
 RL J. Bacteriol. 175:7918-7930(1993).
 CC -1- FUNCTION: REQUIRED FOR REPLICATIVE DNA SYNTHESIS. THIS DNA
 CC POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY (BY
 CC SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
 CC + (DNA)(N).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-C FAMILY. POLC
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U39682; AAC71247.1; -
 CC EMBL: U01807; AAD12336.1; -
 CC EMBL: U01712; AAC43185.1; -
 CC EMBL: U02208; AAD12500.1; -
 CC TIGR: MG031; -
 CC InterPro: IPR000520; Exonuclease.
 CC InterPro: IPR004013; PHP_C.
 CC InterPro: IPR003141; PHP_N.
 CC Pfam: PF00929; Exonuclease; 1.
 CC Pfam: PF02231; PHP_N; 1.
 CC Pfam: PF02811; PHP_C; 1.
 CC SMART: SM00479; EXOIII; 1.
 CC SMART: SM00481; POLI1AC; 1.
 CC TIGRfam: TIGR00573; dnaq; 1.
 CC Transferase: DNA-directed DNA polymerase; DNA replication; Hydrolyase;
 CC Nuclease; Exonuclease; Complete proteome.
 CC DOMAIN 416 583 EXONUCLEASE.
 CC SEQUENCE 1451 AA; 167499 MW; 2DB9E6951F41AAEF CRC64;
 SQ
 Query Match 5.2%; Score 89; DB 1; Length 1451;
 Best Local Similarity 15.9%; Pred. No. 38; Indels 118; Gaps 14;
 Matches 59; Conservative 65; Mismatches 130;
 QY 8 GFS-IGSCFKAAPNPPVLLISASPNKINFTLR-----RKKRFLRYSAVSKFEASALBE 61
 DB 73 GFNEINBEFKADNESE-----KLITFKNEPPFKSKNSSTYTAIKRYFHSFEK 125
 QY 62 TRK-RIVLEPSPHLOEKYSMTGLDGTETLOMLAFKSKIRLLRSM--IENETQVDFEA 118
 DB 126 DKRYKILLIOELSNPNFLSYENDELKACQOTELTEWLIQKSKSFLFMNNAAGFKNFNT 185
 QY 119 GF-----MEPYDPIPCANFTSTNNVIVLDLPHQLDQIDY 159
 DB 186 ALYPDKNESKILKAAVAVSVQYDQKQETKVFATEPIPIHKIMQIDVAVIIQIIELEKTH 245
 QY 160 QDKYNNKIMSIYHKAETFPWGGKLTGESIKFFSPVVMTRPSSSEKHKALFSAFLEY 219
 DB 246 ESLTGKTKLNT---YVDFQLOGSL-----ILKSYTDEKKIEBITIGN----- 286
 QY 220 QAMLEMTIQVREEMEPSSHVRANCEAOKHYLTWRACKDPGHGLIKRLVEAKAKELLRDF- 278
 DB 287 -WIKAHIOV-----ERDPWTQILYGVIREINVEIIPNNK 320
 QY 279 -----LFGVDELGKTFIDFPEYQEDG---TVSKRSI----- 311
 DB 321 RLDSKORRVELVFTKMTAFDGIIND-----IEEYADPAKRGKAKITVTDKINHIYP 374
 QY 312 ---IGSYETR 319
 DB 375 KYEYAAKKYDLK 386
 RESULT 12

APB_CHICK
 ID APB_CHICK STANDARD; PRT; 433 AA.
 AC P11682;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 01-OCT-1989 (Rel. 12, Last annotation update)
 DE Apolipoprotein B (Fragment).
 GN APOB.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88137960; PubMed=3436530;
 RA Kirchgessner T.G., Heinemann C., Svenson K.L., Gordon D.A.,
 RA Nicolsia M., Leberer H.G., Lusis A.J., Williams D.L.;
 RT "Regulation of chicken apolipoprotein B: cloning, tissue
 RT distribution, and estrogen induction of mRNA."
 RL Gene 59:241-251(1987).
 CC -1- FUNCTION: APOLIPOPROTEIN B IS A MAJOR PROTEIN CONSTITUENT OF
 CC CHYLOMICRONS, VLDL AND LDL. IT FUNCTIONS AS A RECOGNITION
 CC SIGNAL FOR THE CELLULAR BINDING AND INTERNALIZATION OF LDL
 CC PARTICLES BY THE APOB/E RECEPTOR.
 CC -1- INDUCTION: WITHIN 24 HOURS AFTER ESTRADIOL ADMINISTRATION,
 CC APOB mRNA IS INCREASED FIVE- TO SEVEN-FOLD IN LIVER BUT IS
 CC UNCHANGED IN INTESTINE AND KIDNEY.
 CC -1- SIMILARITY: WEAK, WITH THE C-TERMINAL 10 % OF HUMAN
 CC APOLIPOPROTEIN B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M18421; AAA48595.1; -
 CC PIR: A29626; A29626.
 CC Plasma; Lipid transport; VLDL; LDL; Chylomicron; Heparin-binding;
 CC Cholesterol metabolism; Glycoprotein; Atherosclerosis.
 CC NON TER 1 1
 CC CARBOHYD 158 158 N-LINKED (GLCNAC...) (POTENTIAL).
 CC SEQUENCE 433 AA; 50847 MW; FD8608C9CFE48925 CRC64;
 SQ
 Query Match 5.1%; Score 88.5; DB 1; Length 433;
 Best Local Similarity 20.6%; Pred. No. 8.9;
 Matches 64; Conservative 50; Mismatches 102; Indels 95; Gaps 16;
 QY 34 TIRRRKRRLVSAVSKFEASALEETRRKRIVLEPSPHLOEKYSMTGLDGTETLOMLA 93
 DB 144 SLAENNVKRYISQIKNFSSQTKLQK--LSEMIQOLVLYIKALIREYEDPTTLGWS----- 194
 QY 94 FSSSKIRLRSAIENETQVDFAGFEMPEYDTPICANFTSTNNVIVLDLNP-LHQ 152
 DB 195 -----VKTYEVEDKVL-----GLNLNMDTLVITNEYAK-----DLSLVTR 232
 QY 153 LTDQI-----DYQDKYNNKIMSIYHKAETFPWGGKLTGESIKFFSPV----- 196
 DB 223 LTDQYRELVEYNTROEYDITDVEGGRQVWELSSAAQDEKIRYMSAVAKRKINEHNRQV 292
 QY 197 -----MWTRESSSEK-----HKAL-----FSAFLEY---YQAMLEMT-----IQ 228
 DB 293 KAKLOEITGOLSDSOEKLNVAKMLIDLTVERYSTFMKATIFELLRFEEDATDSIKPYIA 352
 QY 229 VREEMEPSSHVRANCEAOKHYLTWRACKDPGHGLIKRLVEAKAKELLRDFLNGVDELGT 288
 DB 353 VAE-----GELKIDVPEPMETVINOMPOK--SRBALRKVY-----ELTRALIQGVAYO-GT 399
 QY 289 K-----FTID 293
 RESULT 12

OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxId=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288c / FY73;
 RX MEDLINE=96381248; PubMed=8789261;
 RA Mannhaupt G., Vetter I., Schwarzlose C., Mitzel S., Feldmann H.;
 RT "Analysis of a 26 kb region on the left arm of yeast chromosome XV."
 RL Yeast 12:67-76(1996).
 RN [2]
 RP SEQUENCE OF 548-711 FROM N.A.
 RC STRAIN=5288c;
 RA Ansoerge W., Benes V., Rechmann S., Schwager C., Teodoru C., Voss H.,
 RA Wleemann S.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide from a
 CC peptide comprising four or more residues, with broad specificity.
 CC Also acts on dipeptidyl 2-naphthylamides.
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M49.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: X91067; CAA62529.1; ALT_INIT.
 DR EMBL: 274799; CAA99066.1; -
 DR MEROPS: M49.001; -
 DR SGD: S0005418; YOL057M.
 DR InterPro: IPR005317; Peptidase_M49.
 DR Pfam: PF033571; Peptidase_M49; 1.
 KM Hypothetical protein; Hydrolase; Metalloprotease; Aminopeptidase;
 KM Zinc.
 FT METAL 460 460 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT SITE 461 461 BY SIMILARITY.
 FT METAL 465 465 ZINC (CATALYTIC) (BY SIMILARITY).
 SQ SEQUENCE 711 AA; 80510 MW; C8FEED32834FA683 CRC64;
 Query Match 5.1%; Score 88.5; DB 1; Length 711;
 Best Local Similarity 16.4%; Pred. No. 17; Mismatches 156; Indels 171; Gaps 18;
 Matches 78; Conservative 70;
 QY 18 PNPVLISAŠPKINFTLRKKRRLRLRSVSAVSKFASALEETRRIVLE----- 69
 DB 240 PGGQITLSNAVKVERIFGDHSRE--MRLVASYLKRAQKFAANDQKAMLOEYINNFVIG 297
 QY 70 --PSHLQ-----EKYSMTGLDGTETLOMLAFKSKIRLLNSMAI 107
 DB 298 SSQARKKQAKLWKWDISPVIETNIGFIEFYREPSGIIIG--EFESVIAIONKERTAFSSL 355
 QY 108 ENEETQVDFEAGFMEPEYDPIFCANFTS-----TNVIVVLD----- 146
 DB 356 VNNAEFEISLPMKSK-DYKRPILFNPDPFTSLLEVLTFTGSGIAGINIPYDVRKLTKGR 414
 QY 147 -----LNP-----LHQLTDQ-----TDY 159
 DB 415 NVSLGNILSAAKSSSKHPSPFIQEDRPFEKYQSDSEVOVGHELLGHSGKLLTFE 474
 QY 160 QDKYY-----NKMSIYHYAEFFPWG--GKLTG-----ESTIFF-----SP 194
 DB 475 TDGFNDENENPLGLDGKRVSTYYVGET--WGSKFGQAGPFEBCRAVIAIMFLITNKK 532
 QY 195 LVMMTRFSSSKKHKALFSAFLPEYVOA-----WLEMTIOVREMEPSHVA 240
 DB 533 ILDIGFHDVESQDVIYIAGYIQMARAGLLALEYNNPKTGKMGQPHMARFSIMKTFMKH 592

QY 241 NCEAOKHYLTWRQKPGHGLKRLVGEAKAKELRLDFL-----ENGDELGTGTFID 293
 DB 593 STDKNLKLKLEMNSTNDFAKDKSLIKTAGHCVCADYIKLHLAVYCSGVEGSGKTFID 652
 QY 294 -----YFPEYO--TEDGTASDKRSIIGSKSYETRPMDLTGQFI 328
 DB 653 RSTVTPDLASRLDIIVLSRLRPRQFIQSNYSYIDNNKRVTLKXEYDEPQGMAGSFL 707
 RESULT 15
 ID ODOI_RICPR STANDARD; PRT; 936 AA.
 AC 092DY3;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 2-oxoglutarate dehydrogenase E1 component (EC 1.2.4.2) (Alpha-
 GN SUCA OR Rp180.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiidae; Rickettsia.
 OX NCBI_TaxId=782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=99039495; PubMed=9823893;
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Sichteritz-Ponten T., Alismark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria."
 RL Nature 396:133-140(1998).
 CC -1- FUNCTION: THE 2-OXOGLUTARATE DEHYDROGENASE COMPLEX CATALYZES THE
 CC OVERALL CONVERSION OF 2-OXOGLUTARATE TO SUCCINYL-COA & CO(2). IT
 CC CONTAINS MULTIPLE COPIES OF 3 ENZYMAIC COMPONENTS: 2-OXOGLUTARATE
 CC DEHYDROGENASE (E1), DIHYDROLIPOAMIDE SUCCINYLTRANSFERASE (E2) AND
 CC LIPOAMIDE DEHYDROGENASE (E3) (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: 2-oxoglutarate + lipamide - S-
 CC succinylidihydroliopamide + CO(2).
 CC -1- COFACTOR: THIAMINE PYROPHOSPHATE (BY SIMILARITY).
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: AJ235270; CAA16447.1; -
 DR InterPro: IPR001017; Dehydrogenase_E1.
 DR Pfam: PF00676; E1_denydrog; 1.
 DR Pfam: PF02779; transket_pyr; 1.
 DR TIGRPFAMs: TIGR00239; 2oxo_dh_E1; 1.
 KM Glycyls: Oxidoreductase; Flavoprotein; Thiamine pyrophosphate;
 KM Complete proteome.
 SQ SEQUENCE 936 AA; 105795 MW; 30EA24895B4758AC CRC64;
 Query Match 5.1%; Score 88.5; DB 1; Length 936;
 Best Local Similarity 22.4%; Pred. No. 24;
 Matches 77; Conservative 49; Mismatches 129; Indels 89; Gaps 20;
 QY 39 KRFLRLRSVSAVSKFASALEETRRRIYLPESHQEKY-----SMTGLDGTKE 88
 DB 500 KEKFRITLDQ-EYEQ-AKSYKQET-----HPEGYWKGISIRIGKDATGVVKKI- 547
 QY 89 LQMLAFKSKI-----RLRSMALNEMTM--QVDFPAGFMEPEYDPIFCANFTS 137
 DB 548 LODLGTIKCEIPKRDPAINKRLIRLFEVRKTTLTTPDIPIDMAVAEOLAF-AHLICSGI--- 603

```

QY 138 FNWNLVAVLDL-----NPLHOLTDORUDY-----ODKYKIMSIYKAYETP 179
Db 604 -NIRLTGODSANGTSHRSILHANOQIDTTYIPLNLSKTQAKYEYANSWLSEYALGFE 662
QY 180 WGGKLTGESIKFFSPLVM-TRESSSEKHKALFSALEYQA-WLEMTIOW-----RE 231
Db 663 YGXSILNP-----KNLVLWEAQGFDPANQAQIIFDPFISSATKMWLRMSGLVLLPHAFE 717
QY 232 EMBEBSHVANCE-----AQHKYLTWRQKQDPGHLLKRLVGEAKAKELL----- 275
Db 718 GGGPHESSSRRLERFIOLAEEKNNTIYPTTPASIFLRLRQILLESSTRKPLTVMSPKSLR 777
QY 276 RDLFENGVDLGS-TKFEIDYFPEYQTEGDSVSDKRSII--GKSY 316
Db 778 HKYAVSKDELCEGNTTFFIPILDEV--RKIDTNNTKTYILSGKYY 820

```

Search completed: June 24, 2003, 18:37:11
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 18:53:01 ; Search time 16 Seconds

(Without alignments)
605.009 Million cell updates/sec

Title: US-09-870-406a-33

Perfect score: 329
Sequence: 1 MALSMFGFSIGSCFKAPNP.....SIIGKSYETRPMDLTGQFIG 329

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match	Length	DB	ID	Description
1	7	2.1	11	4	US-09-425-638A-69	Sequence 69, Appl
2	7	2.1	11	4	US-09-543-004-69	Sequence 69, Appl
3	7	2.1	109	4	US-09-425-638A-22	Sequence 22, Appl
4	7	2.1	109	4	US-09-543-004-22	Sequence 22, Appl
5	7	2.1	398	2	US-08-599-171A-29	Sequence 29, Appl
6	7	2.1	398	2	US-08-646-590B-29	Sequence 29, Appl
7	7	2.1	398	3	US-09-069-226-29	Sequence 29, Appl
8	7	2.1	398	4	US-09-412-184-29	Sequence 29, Appl
9	7	2.1	830	4	US-09-562-737-31	Sequence 31, Appl
10	7	2.1	830	4	US-09-562-737-36	Sequence 36, Appl
11	7	2.1	830	4	US-09-562-737-37	Sequence 37, Appl
12	7	2.1	830	4	US-09-562-737-38	Sequence 38, Appl
13	7	2.1	940	2	US-08-938-365-4	Sequence 4, Appl
14	7	2.1	941	1	US-08-343-760A-2	Sequence 12, Appl
15	7	2.1	1142	1	US-08-904-871-12	Sequence 12, Appl
16	6	1.8	8	1	US-08-305-172B-15	Sequence 15, Appl
17	6	1.8	11	4	US-09-425-638A-68	Sequence 68, Appl
18	6	1.8	11	4	US-09-543-004-68	Sequence 68, Appl
19	6	1.8	14	4	US-08-973-131-75	Sequence 75, Appl
20	6	1.8	16	2	US-09-133-774-4	Sequence 4, Appl
21	6	1.8	16	3	US-09-303-862-4	Sequence 4, Appl
22	6	1.8	17	4	US-09-391-104-14	Sequence 14, Appl
23	6	1.8	22	5	PCT-US97-08094-5	Sequence 5, Appl
24	6	1.8	38	1	US-07-977-630-66	Sequence 66, Appl
25	6	1.8	64	1	US-08-344-695-14	Sequence 14, Appl
26	6	1.8	64	4	US-09-627-376-6	Sequence 6, Appl
27	6	1.8	71	3	US-08-297-395-8	Sequence 8, Appl

28	6	1.8	76	1	US-08-848-252-4	Sequence 4, Appl
29	6	1.8	86	4	US-09-134-001C-5516	Sequence 5516, Ap
30	6	1.8	93	2	US-08-341-843B-32	Sequence 32, Appl
31	6	1.8	93	2	US-08-427-497E-37	Sequence 37, Appl
32	6	1.8	109	4	US-09-425-638A-20	Sequence 20, Appl
33	6	1.8	109	4	US-09-425-638A-21	Sequence 21, Appl
34	6	1.8	109	4	US-09-425-638A-39	Sequence 39, Appl
35	6	1.8	109	4	US-09-425-638A-40	Sequence 40, Appl
36	6	1.8	109	4	US-09-425-638A-41	Sequence 41, Appl
37	6	1.8	109	4	US-09-425-638A-42	Sequence 42, Appl
38	6	1.8	109	4	US-09-425-638A-43	Sequence 43, Appl
39	6	1.8	109	4	US-09-425-638A-44	Sequence 44, Appl
40	6	1.8	109	4	US-09-425-638A-45	Sequence 45, Appl
41	6	1.8	109	4	US-09-543-004-20	Sequence 20, Appl
42	6	1.8	109	4	US-09-543-004-21	Sequence 21, Appl
43	6	1.8	109	4	US-09-543-004-39	Sequence 39, Appl
44	6	1.8	109	4	US-09-543-004-40	Sequence 40, Appl
45	6	1.8	109	4	US-09-543-004-41	Sequence 41, Appl

ALIGNMENTS

RESULT 1
US-09-425-638A-69
Sequence 69, Application US/09425638A
Patent No. 6342587
GENERAL INFORMATION:
APPLICANT: Carlos F. Barbas III, Christoph Rader, Gerd Ritter, Sydney Welt an
TITLE OF INVENTION: A33 ANTIGEN SPECIFIC IMMUNOGLOBULIN PRODUCTS AND USES THE
FILE REFERENCE: LUD 5630
CURRENT APPLICATION NUMBER: US/09/425,638A
CURRENT FILING DATE: 1999-10-22
NUMBER OF SEQ ID NOS: 129
SEQ ID NO 69
LENGTH: 11
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
US-09-425-638A-69

Query Match
Best Local Similarity 100.0%; DB 4; Length 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 277 DLENV 283
Db 4 DLENV 10

RESULT 2
US-09-543-004-69
Sequence 69, Application US/09543004
Patent No. 6346249
GENERAL INFORMATION:
APPLICANT: Carlos F. Barbas III, Christoph Rader, Gerd Ritter, Sydney Welt an
TITLE OF INVENTION: A33 ANTIGEN SPECIFIC IMMUNOGLOBULIN PRODUCTS AND USES THE
FILE REFERENCE: LUD 5630.1
CURRENT APPLICATION NUMBER: US/09/543,004
CURRENT FILING DATE: 2000-04-04
PRIOR APPLICATION NUMBER: 09/425,638
PRIOR FILING DATE: 1999-10-22
NUMBER OF SEQ ID NOS: 129
SEQ ID NO 69
LENGTH: 11
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
US-09-543-004-69

Query Match 2.1%; Score 7; DB 4; Length 11;

Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 277 DFLFNGV 283
Db 4 DFLFNGV 10

RESULT 3

US-09-425-638A-22
; Sequence 22, Application US/09425638A
; Patent No. 6342587
; GENERAL INFORMATION:
; APPLICANT: Carlos F. Barbas III, Christoph Rader, Gerd Ritter, Sydney Welt and
; APPLICANT: Lloyd J. Old
; TITLE OF INVENTION: A33 ANTIGEN SPECIFIC IMMUNOGLOBULIN PRODUCTS AND USES THEREO
; FILE REFERENCE: LUD 5630
; CURRENT APPLICATION NUMBER: US/09/425,638A
; CURRENT FILING DATE: 1999-10-22
; NUMBER OF SEQ ID NOS: 129
; SEQ ID NO 22
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
US-09-425-638A-22

Query Match 2.1%; Score 7; DB 4; Length 109;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 277 DFLFNGV 283
Db 27 DFLFNGV 33

RESULT 4

US-09-543-004-22
; Sequence 22, Application US/09543004
; Patent No. 6346249
; GENERAL INFORMATION:
; APPLICANT: Carlos F. Barbas III, Christoph Rader, Gerd Ritter, Sydney Welt and
; APPLICANT: Lloyd J. Old
; TITLE OF INVENTION: A33 ANTIGEN SPECIFIC IMMUNOGLOBULIN PRODUCTS AND USES THEREO
; FILE REFERENCE: LUD 5630.1
; CURRENT APPLICATION NUMBER: US/09/543,004
; CURRENT FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: 09/425,638
; PRIOR FILING DATE: 1999-10-22
; NUMBER OF SEQ ID NOS: 129
; SEQ ID NO 22
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
US-09-543-004-22

Query Match 2.1%; Score 7; DB 4; Length 109;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 277 DFLFNGV 283
Db 27 DFLFNGV 33

RESULT 5

US-08-599-171A-29
; Sequence 29, Application US/08599171A
; Patent No. 5814473
; GENERAL INFORMATION:
; APPLICANT: WARREN, Patrick V.
; TITLE OF INVENTION: TRANSMINASES AND AMINOTRANSFERASES

NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:

ADDRESSEE: CARELLA, BYRNE, BAIN, GIUFFILIAN,
ADDRESSEE: CECCHI, STEWART & OLSPREIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH DISKETTE

COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS

SOFTWARE: WORD PERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/599,171A

FILING DATE: Concurrently

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: HERRON, CHARLES J.

REGISTRATION NUMBER: 28,019

REFERENCE/DOCKET NUMBER: 331400-38

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-994-1700

TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 29:

SEQUENCE CHARACTERISTICS:

LENGTH: 398 AMINO ACIDS

TYPE: AMINO ACID

TOPOLOGY: LINEAR

MOLECULE TYPE: PROTEIN

US-08-599-171A-29

Query Match 2.1%; Score 7; DB 2; Length 398;

Best Local Similarity 100.0%; Pred. No. 79;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 270 KAKELLR 276
Db 22 KAKELLR 28

RESULT 6

US-08-646-590B-29
; Sequence 29, Application US/08646590B
; Patent No. 5962283
; GENERAL INFORMATION:
; APPLICANT: Warren, Patrick V.
; APPLICANT: Swanson, Ronald V.
; TITLE OF INVENTION: TRANSMINASES AND AMINOTRANSFERASES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,590B
; FILING DATE: 08-May-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/599,171

FILING DATE: 09-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/01094
FILING DATE: 21-January-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Ph.D., Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/017001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
FAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 398 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-646-590b-29

Query Match 2.1%; Score 7; DB 2; Length 398;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 270 KAKELLR 276
Db 22 KAKELLR 28

RESULT 7
US-09-069-226-29
Sequence 29, Application US/09069226
Patent No. 6013509
GENERAL INFORMATION:
APPLICANT: WARREN, Patrick V.
TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLIAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/069,226
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/599,171
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: HERRON, CHARLES J.
REGISTRATION NUMBER: 28,019
REFERENCE/DOCKET NUMBER: 331400-38
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
FAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 398 AMINO ACIDS
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-09-069-226-29

Query Match 2.1%; Score 7; DB 3; Length 398;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 270 KAKELLR 276
Db 22 KAKELLR 28

RESULT 8
US-09-412-184-29
Sequence 29, Application US/09412184
Patent No. 6268188
GENERAL INFORMATION:
APPLICANT: Warren, Patrick V.
APPLICANT: Swanson, Ronald V.
TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: US
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/412,184
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/646,590
FILING DATE: 08-May-1996
APPLICATION NUMBER: 08/599,171
FILING DATE: 09-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/01094
FILING DATE: 21-January-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Ph.D., Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/017001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5099
FAX: 619/678-5070
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 398 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-09-412-184-29

Query Match 2.1%; Score 7; DB 4; Length 398;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 270 KAKELLR 276
Db 22 KAKELLR 28

RESULT 9
US-09-562-737-31
Sequence 31, Application US/09562737
Patent No. 6428967
GENERAL INFORMATION:

APPLICANT: Herz, Joachim
APPLICANT: Gotthardt, Michael
TITLE OF INVENTION: LDL Receptor Signaling Pathways
FILE REFERENCE: UTSW0708
CURRENT APPLICATION NUMBER: US/09/562,737
CURRENT FILING DATE: 2000-05-01
NUMBER OF SEQ ID NOS: 132
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 31
LENGTH: 830
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-562-737-31

Query Match 2.1%; Score 7; DB 4; Length 830;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 214 AFLEYQ 220
DB 809 AFLEYQ 815

RESULT 10
US-09-562-737-36
Sequence 36, Application US/09562737
Patent No. 6428967
GENERAL INFORMATION:
APPLICANT: Herz, Joachim
APPLICANT: Gotthardt, Michael
TITLE OF INVENTION: LDL Receptor Signaling Pathways
FILE REFERENCE: UTSW0708
CURRENT APPLICATION NUMBER: US/09/562,737
CURRENT FILING DATE: 2000-05-01
NUMBER OF SEQ ID NOS: 132
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 36
LENGTH: 830
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-562-737-36

Query Match 2.1%; Score 7; DB 4; Length 830;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 214 AFLEYQ 220
DB 809 AFLEYQ 815

RESULT 11
US-09-562-737-37
Sequence 37, Application US/09562737
Patent No. 6428967
GENERAL INFORMATION:
APPLICANT: Herz, Joachim
APPLICANT: Gotthardt, Michael
TITLE OF INVENTION: LDL Receptor Signaling Pathways
FILE REFERENCE: UTSW0708
CURRENT APPLICATION NUMBER: US/09/562,737
CURRENT FILING DATE: 2000-05-01
NUMBER OF SEQ ID NOS: 132
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 37
LENGTH: 830
TYPE: PRT

ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-562-737-37

Query Match 2.1%; Score 7; DB 4; Length 830;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 214 AFLEYQ 220
DB 809 AFLEYQ 815

RESULT 12
US-09-562-737-38
Sequence 38, Application US/09562737
Patent No. 6428967
GENERAL INFORMATION:
APPLICANT: Herz, Joachim
APPLICANT: Gotthardt, Michael
TITLE OF INVENTION: LDL Receptor Signaling Pathways
FILE REFERENCE: UTSW0708
CURRENT APPLICATION NUMBER: US/09/562,737
CURRENT FILING DATE: 2000-05-01
NUMBER OF SEQ ID NOS: 132
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 38
LENGTH: 830
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-562-737-38

Query Match 2.1%; Score 7; DB 4; Length 830;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 214 AFLEYQ 220
DB 809 AFLEYQ 815

RESULT 13
US-08-938-365-4
Sequence 4, Application US/08938365
Patent No. 5989909
GENERAL INFORMATION:
APPLICANT: Yang, Pan
TITLE OF INVENTION: HUCHORDIN AND USRS THEREOF
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,365
FILING DATE: 26-SEP-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Melkielejohn, Ph.D., Anita L.
REGISTRATION NUMBER: 35,283
REFERENCE/DOCKET NUMBER: 09404/040001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 940 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-938-365-4

Query Match 2.1%; Score 7; DB 2; Length 940;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 KHKALFS 213
DB 275 KHKALFS 281

RESULT 14
US-08-343-760A-2
Sequence 2, Application US/08343760A
Patent No. 5679783
GENERAL INFORMATION:
APPLICANT: De Robertis, Edward M
TITLE OF INVENTION: Tissue Differentiation Affecting
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Majestic, Parsons, Siebert & Hsue
STREET: Four Embarcadero Center, Suite 1450
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,760A
FILING DATE: 22-Nov-1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Siebert, J. Suzanne
REGISTRATION NUMBER: 28,758
REFERENCE/DOCKET NUMBER: 3100.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 363-5556
TELEFAX: (415) 362-5418
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 941 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-343-760A-2

Query Match 2.1%; Score 7; DB 1; Length 941;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 KHKALFS 213
DB 275 KHKALFS 281

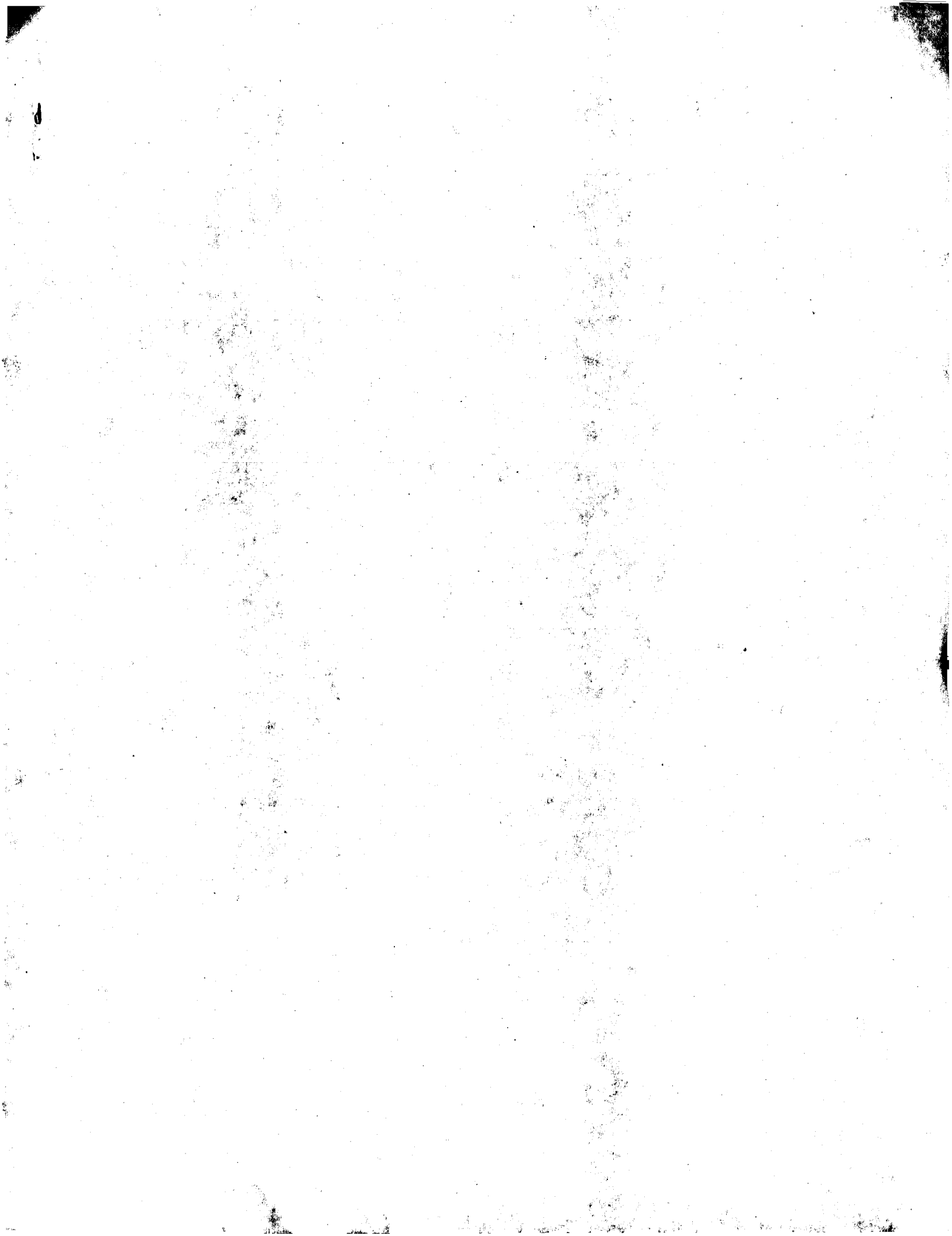
DB 276 KHKALFS 282

RESULT 15
US-08-904-871-12
Sequence 12, Application US/08904871
Patent No. 6046014
GENERAL INFORMATION:
APPLICANT: Lagarias, John C
TITLE OF INVENTION: PHYTOFLUORS AS FLUORESCENT LABELS
FILE REFERENCE: 2500.134050 DC OTT Lagarias Patent
CURRENT APPLICATION NUMBER: US/08/904,871
EARLIER FILING DATE: 1997-08-01
EARLIER APPLICATION NUMBER: 60/023,217
NUMBER OF SEQ ID NOS: 16
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 12
LENGTH: 1142
TYPE: PRT
ORGANISM: Alga (Mesotaelinum)
US-08-904-871-12

Query Match 2.1%; Score 7; DB 3; Length 1142;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 268 EAKAKEL 274
DB 905 EAKAKEL 911

Search completed: June 24, 2003, 18:56:16
Job time : 16 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 17:17:40 ; Search time 44 Seconds
(without alignments)
996.351 Million cell updates/sec

Title: US-09-870-406A-33

Perfect score: 1724

Sequence: 1 MALSMERGFSGSCFKAPNP.....SIICKSYETRWDLTGQFIG 329

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1703	98.8	329	23	AAM50863
2	1698	98.5	327	21	AAG18223
3	1681	97.5	323	21	AAG18224
4	1330	77.1	250	21	AAG18225
5	110.5	6.4	514	16	AA878269
6	110.5	6.4	515	16	AA872449
7	110.5	6.4	515	18	AAW31406
8	110.5	6.4	515	21	AA97547
9	110.5	6.4	515	22	AA97547
10	110.5	6.4	515	23	AB876588

11	110.5	6.4	515	23	AB806935
12	110.5	6.4	515	23	AB847852
13	110.5	6.4	515	23	AAU12151
14	105	6.1	548	19	AAW37444
15	105	6.1	548	20	AAV01585
16	105	6.1	548	22	AA65879
17	104.5	6.1	549	8	AA870338
18	94.5	5.5	1886	22	AA879823
19	93.5	5.4	386	23	AB853527
20	92	5.3	1348	22	AB860761
21	91	5.3	830	22	AB865581
22	90.5	5.2	727	23	ABP39222
23	90.5	5.2	952	22	AA879089
24	90.5	5.2	1046	23	AB854658
25	90	5.2	501	22	AB860014
26	90	5.2	572	8	AA870579
27	89.5	5.2	775	22	AB893302
28	89.5	5.2	855	16	AA872855
29	89.5	5.2	855	16	AA872855
30	89	5.2	369	15	AA862443
31	88.5	5.1	512	20	AAV07388
32	88.5	5.1	514	17	AAW12955
33	88.5	5.1	514	20	AAV15417
34	88.5	5.1	514	20	AAV07383
35	88.5	5.1	514	21	AAV99604
36	88.5	5.1	515	18	AAW31502
37	88.5	5.1	515	19	AAW48263
38	88.5	5.1	515	20	AAV25153
39	88.5	5.1	532	23	AB890169
40	88.5	5.1	559	8	AA870580
41	88.5	5.1	573	23	ABP39968
42	88.5	5.1	1156	22	AA893207
43	88.5	5.1	1884	22	AA863241
44	88.5	5.1	1885	22	AA878839
45	88	5.1	359	23	AA880302

ALIGNMENTS

RESULT 1	
AAM50863	
ID	AAM50863 standard; Protein; 329 AA.
AC	AAM50863;
DT	07-MAY-2002 (first entry)
XX	
DE	Arabidopsis phytochromobilin synthase HY2.
KW	HY2, biliverdin reductase; phytochromobilin synthase;
KW	phytochrome; phytofluor; plant; enzyme.
OS	Arabidopsis thaliana.
XX	
FT	Key
FT	Peptide
FT	Protein
FT	Location/Qualifiers
FT	1..45
FT	/label= Translt_peptide
FT	46..329
FT	/label= Mature_protein
PN	WO200194548-A2.
XX	
PD	13-DEC-2001.
XX	
PF	05-JUN-2001; 2001WO-US18326.
XX	
PR	08-JUN-2000; 2000US-210286P.
PR	26-FEB-2001; 2001US-271758P.
XX	
PR	29-MAY-2001; 2001US-0210286.
XX	
PA	(RECC) UNIV CALIFORNIA.
XX	

B. steatothermophi
Bacillus alpha amy
Bacillus TERMAHL-
B. steatothermophi
An alpha-amyase (
B. steatothermophi
Alpha-amyase gene
Human protein SEQ
Lactococcus lactis
Drosophila melanog
Staphylococcus epi
Human protein sequ
Lactococcus lactis
Drosophila melanog
Human wound healin
Plasmodium falcipar
Human protein sequ
A. nidulans PKSA,
R. CytII encoded by
Mutant termamyl(RT
Alpha-amyase. Ba
Bacillus steatother
Wild type Termamyl
Bacillus steatother
Bacillus sp. alpha
Bacillus sp. alpha
Bacillus sp. alpha
Human polyprotein
Plasmodium falcipar
Staphylococcus epi
Human protein sequ
Human breast cance
Human protein SEQ
C. albicans hyphal

PI Lagarias JC, Kochi T, Frankenberg N, Gambetta GA, Montgomery BL;
XX WPI: 2002-195566/25.
DR N-PSDB: ABA91766.
XX
PT Novel isolated HY2 family bilin reductase having bilin reductase
PT activity, useful for converting biliverdin to phytylbilin, and for
PT producing a photoactive holophytochrome and/or phytyfluor
XX
XX
PS Example 3; Fig 3B; 102pp; English.
XX

CC The present sequence is that of the HY2 protein of Arabidopsis
CC thaliana ecotype Columbia, as predicted from cDNA sequence
CC analysis. HY2 is a ferredoxin-dependent biliverdin reductase that
CC has phytylholobilin synthase activity. The 329-amino acid protein
CC has a predicted molecular mass of 38.1 kDa. The first 45 N-terminal
CC amino acid residues are predicted to form a chloroplast transit
CC peptide, suggesting that the HY2 protein is localised in the
CC chloroplast. HY2 is related to a family of proteins found in
CC oxygenic photosynthetic bacteria. It is an example of bilin
CC reductases of the invention, which are useful e.g. for the
CC conversion of biliverdin to phytylbilin and the assembly of
CC holophytochromes or phytyfluors. A claimed method of producing
CC a photoactive holophytochrome involves co-expressing a haem
CC oxygenase, an apophytochrome and a ferredoxin-dependent bilin
CC reductase, in a cell, where the cell produces the photoactive
CC holophytochrome and, where the apophytochrome and/or the
CC bilin reductase are expressed by heterologous nucleic acids.
CC The cell may be an algal, yeast, bacterial, plant, insect or
CC mammalian cell, and the bilin reductase is preferably an HY2
CC family bilin reductase.
XX

SO Sequence 329 AA;

Query Match 98.8%; Score 1703; DB 23; Length 329;
Best Local Similarity 99.1%; Pred. No. 1,1e-166;
Matches 326; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAISMEFGESIGCFKAPNPVLLISAPNKINFTLRKRKFLRLRSVAVSYKKEFAESALE 60
DB 1 MAISMEFGESIGCFKAPNPVLLISAPNKINFTLRKRKFLRLRSVAVSYKKEFAESALE 60
QY 61 ETRKRIYLEPSHIOEKSSMTGIDGKTELQMLAFKSKIKLMSALENTMOVFPFAGF 120
DB 61 ETRKRIYLEPSHIOEKSSMTGIDGKTELQMLAFKSKIKLMSALENTMOVFPFAGF 120
QY 121 MEPEYDTPICANFFSTNNVNIYVLDLPLHQLTDQTDYODKYYNKIMSIYHKAETFPW 180
DB 121 MEPEYDTPICANFFSTNNVNIYVLDLPLHQLTDQTDYODKYYNKIMSIYHKAETFPW 180
QY 181 GGLTGESITKFEPSLWMTFFSSSKKKRALEFSAFLFYQAWLEMTIOVREEMPSHVRA 240
DB 181 GGLTGESITKFEPSLWMTFFSSSKKKRALEFSAFLFYQAWLEMTIOVREEMPSHVRA 240
QY 241 NCAOQKRYLWRAQKPGHGLKRLVGEAKAKELLDLFLNGVDLGTKEFDYFEXOT 300
DB 241 NCAOQKRYLWRAQKPGHGLKRLVGEAKAKELLDLFLNGVDLGTKEFDYFEXOT 300
QY 301 EDGTVDKRSIIIGKSYETRPWDLTGQFIG 329
DB 301 EDGTVDKRSIIIGKSYETRPWDLTGQFIG 329

RESULT 2
AAG18223
ID AAG18223 standard; Protein; 327 AA.

XX AAG18223;
AC AAG18223;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 19545.
XX

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129645.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 07-MAY-1999; 99US-0132486.
PR 11-MAY-1999; 99US-0132487.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134841.
PR 20-MAY-1999; 99US-0135324.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.

DB 299 EDGTVSDKRSIIKSYETRPWDLTGQFIG 327

RESULT 3

AA18224 standard; Protein; 323 AA.

AA18224;

17-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 19546.

Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 9905-0121825.

05-MAR-1999; 9905-0123180.

09-MAR-1999; 9905-0123548.

23-MAR-1999; 9905-0125788.

25-MAR-1999; 9905-0126264.

29-MAR-1999; 9905-0126785.

01-APR-1999; 9905-0127462.

06-APR-1999; 9905-0128234.

16-APR-1999; 9905-0128714.

19-APR-1999; 9905-0129845.

21-APR-1999; 9905-0130077.

23-APR-1999; 9905-0130449.

28-APR-1999; 9905-0130891.

30-APR-1999; 9905-0131449.

07-MAY-1999; 9905-0132048.

04-MAY-1999; 9905-0132484.

05-MAY-1999; 9905-0132485.

06-MAY-1999; 9905-0132486.

07-MAY-1999; 9905-0132487.

11-MAY-1999; 9905-0132863.

14-MAY-1999; 9905-0134218.

14-MAY-1999; 9905-0134219.

14-MAY-1999; 9905-0134221.

14-MAY-1999; 9905-0134370.

18-MAY-1999; 9905-0134768.

19-MAY-1999; 9905-0134941.

20-MAY-1999; 9905-0135124.

21-MAY-1999; 9905-0135353.

24-MAY-1999; 9905-0135629.

25-MAY-1999; 9905-0136021.

27-MAY-1999; 9905-0136392.

28-MAY-1999; 9905-0136782.

01-JUN-1999; 9905-0137222.

03-JUN-1999; 9905-0137528.

04-JUN-1999; 9905-0137502.

07-JUN-1999; 9905-0137724.

08-JUN-1999; 9905-0138094.

10-JUN-1999; 9905-0138540.

10-JUN-1999; 9905-0138847.

14-JUN-1999; 9905-0139119.

16-JUN-1999; 9905-0139452.

16-JUN-1999; 9905-0139453.

17-JUN-1999; 9905-0139492.

18-JUN-1999; 9905-0139454.

18-JUN-1999; 9905-0139455.

18-JUN-1999; 9905-0139456.

18-JUN-1999; 9905-0139457.

18-JUN-1999; 9905-0139458.

18-JUN-1999; 9905-0139459.

18-JUN-1999; 9905-0139460.

18-JUN-1999; 9905-0139461.

18-JUN-1999; 9905-0139462.

18-JUN-1999; 9905-0139463.

18-JUN-1999; 9905-0139750.

18-JUN-1999; 9905-0139763.

21-JUN-1999; 9905-0139817.

22-JUN-1999; 9905-0139899.

23-JUN-1999; 9905-0140353.

23-JUN-1999; 9905-0140354.

24-JUN-1999; 9905-0140695.

28-JUN-1999; 9905-0140823.

29-JUN-1999; 9905-0140981.

30-JUN-1999; 9905-0141287.

01-JUL-1999; 9905-0141842.

01-JUL-1999; 9905-0142154.

02-JUL-1999; 9905-0142055.

06-JUL-1999; 9905-0142390.

08-JUL-1999; 9905-0142803.

09-JUL-1999; 9905-0142920.

12-JUL-1999; 9905-0142977.

13-JUL-1999; 9905-0143542.

14-JUL-1999; 9905-0143624.

15-JUL-1999; 9905-0144005.

16-JUL-1999; 9905-0144085.

16-JUL-1999; 9905-0144086.

19-JUL-1999; 9905-0144325.

19-JUL-1999; 9905-0144331.

19-JUL-1999; 9905-0144332.

19-JUL-1999; 9905-0144333.

19-JUL-1999; 9905-0144334.

19-JUL-1999; 9905-0144335.

20-JUL-1999; 9905-0144352.

20-JUL-1999; 9905-0144632.

20-JUL-1999; 9905-0144684.

21-JUL-1999; 9905-0144814.

21-JUL-1999; 9905-0145086.

21-JUL-1999; 9905-0145088.

22-JUL-1999; 9905-0145085.

22-JUL-1999; 9905-0145087.

22-JUL-1999; 9905-0145089.

22-JUL-1999; 9905-0145192.

23-JUL-1999; 9905-0145145.

23-JUL-1999; 9905-0145218.

23-JUL-1999; 9905-0145224.

26-JUL-1999; 9905-0145276.

27-JUL-1999; 9905-0145913.

27-JUL-1999; 9905-0145918.

27-JUL-1999; 9905-0145919.

28-JUL-1999; 9905-0145951.

02-AUG-1999; 9905-0146386.

02-AUG-1999; 9905-0146388.

02-AUG-1999; 9905-0146389.

03-AUG-1999; 9905-0147038.

04-AUG-1999; 9905-0147204.

04-AUG-1999; 9905-0147302.

05-AUG-1999; 9905-0147192.

05-AUG-1999; 9905-0147260.

06-AUG-1999; 9905-0147303.

06-AUG-1999; 9905-0147416.

09-AUG-1999; 9905-0147493.

09-AUG-1999; 9905-0147935.

10-AUG-1999; 9905-0148171.

11-AUG-1999; 9905-0148319.

12-AUG-1999; 9905-0148341.

13-AUG-1999; 9905-0148565.

13-AUG-1999; 9905-0148684.

16-AUG-1999; 9905-0149368.

17-AUG-1999; 9905-0149175.

```
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162143.
PR 29-OCT-1999; 99US-0162144.

Query Match 97.5% Score 1681; DB 21; Length 323;
Best Local Similarity 99.1%; Pred. No. 2e-164;
Matches 322; Conservative 0; Mismatches 1; Indels 2; Gaps 1;
```

```
OY 5 MEFGRSIGCFKADNPVLISASPNKINFTLRKKRKLRLVSAVSYPEFAESALEETRK 64
DB 1 MEFGRSIGCFKADNPVLISASPNKINFTLRKKRKLRLVSAVSYPEFAESALEETRK 60
OY 65 RIVLEPSHLQEKYSMTGLDKTELQMLAFKSSKIRLLRSMAIENETQVDFDAGFMEPE 124
DB 61 RIVLEPSHLQ--YSSMTGLDKTELQMLAFKSSKIRLLRSMAIENETQVDFDAGFMEPE 118
OY 125 YDFPIFCANFTSTNVNIIVLDLNPJLHQLTQDTQDYQDKYKINSIYHKVAFETFPWGCKL 184
```

```
DB 119 YDFPIFCANFTSTNVNIIVLDLNPJLHQLTQDTQDYQDKYKINSIYHKVAFETFPWGCKL 178
OY 185 TGESIKFFSPVLMYMTTRSSSEKHKALFSALLEYIOAMLEMTIOYREEMESHVANCEA 244
DB 179 TGESIKFFSPVLMYMTTRSSSEKHKALFSALLEYIOAMLEMTIOYREEMESHVANCEA 238
OY 245 QHKYLTWPAOKDPGGLIKRLVGEAKAKELLRDPLFNGVDELGTGRTFIDYFPEYOTEDGT 304
DB 239 QHKYLTWPAOKDPGGLIKRLVGEAKAKELLRDPLFNGVDELGTGRTFIDYFPEYOTEDGT 298
OY 305 VSDKRSIIGKSYETRPMDLTGQFTG 329
DB 299 VSDKRSIIGKSYETRPMDLTGQFTG 323

RESULT 4
AAG18225
ID AAG18225 standard; Protein; 250 AA.
AC AAG18225;
XX
XX 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 19547.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134376.
PR 20-MAY-1999; 99US-0134924.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
```

PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 17-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142910.
PR 12-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156658.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

XX (NOVO) NOVO-NORDISK AS.
 PA Svendsen A, Kjaerulff S, Bisgard-Frantzen H, Andersen C;
 PI WPI: 2000-387777/33.
 DR N-PSDB: AAA48482.
 XX
 PT Variant of parent termamyl-like alpha amylase useful for washing,
 PT textile desizing and starch liquefaction, comprising alterations in one
 PT or more solvent exposed amino acid residues
 PS
 XX Claim 8: Fig 1: 80pp: English.
 XX
 CC The present sequence is a parent alpha-amylase from which mutants
 CC with increased stability at acidic pH, low calcium concentration and
 CC high temperatures have been derived. A variant may contain mutations in
 CC one or more solvent exposed amino acid residues to increase the overall
 CC hydrophobicity of the enzyme or the overall number of methyl groups in
 CC the side chains of exposed residues may be increased. The mutations can
 CC be incorporated by site-directed mutagenesis or by random mutagenesis.
 CC As a result of their increased stability, the variants are suitable for
 CC the industrial processing of starch, i.e. starch liquefaction and
 CC saccharification. They may also be useful for washing, dishwashing and
 CC textile desizing. Hybrid alpha-amylases comprising partial amino acid
 CC sequences derived from two or more alpha-amylases have also been created
 CC in order to increase enzyme stability.
 CC Note: According to the specification, the present sequence and the
 CC sequence shown in AA959604 are the same.
 CC
 XX
 SQ Sequence 515 AA:

Query Match 6.4%; Score 110.5; DB 21; Length 515;
 Best Local Similarity 19.4%; Pred. No. 0.028; Mismatches 120; Indels 113; Gaps 16;
 Matches 68; Conservative 50; Mismatches 120; Indels 113; Gaps 16;
 QY 50 SYKEFAESALEETRRRIYLEPHSHLOEKYSMTGLDQKTELQMLAFKSKIRLLRSMAIEN 109
 DB 195 NYDIYMTADLDMDHDEVTTELKNMGKWTNTNIDG-----FRIDAVKH 238
 QY 110 ETMOVF-DPAGMEPEYDPIFCANFETSTNVIYVLDNPLHQLTDQD-----Y 159
 DB 239 IKFSFPDMLSYRSQTGKPLFTVGEYWS-----YDINKLHNTTKDGTMSLFDAPL 291
 QY 160 QDKYY-----NKIMSIYHKAETP-----PWGKRLTG 186
 DB 292 HNKFTYASKSGAFDMRTLMNTLMKDQPTLAVTFVDNHDTEPGALQSMVDPW----- 345
 QY 187 ESIKFSPPLVMTTRSSSEKHKALFSAFLEYQAMLEMTIQVREMEPSHVRANCEA-- 244
 DB 346 -----FKPLA-YAFILTRQEGYPCVF--YGDYIGIPQVNIPIPSLKSIDPLILARRDYAYG 397
 QY 245 -QHKYLT-----WRAQ--KDPGHGLKRL-----VGEAKAKELLRLDFLFCV 283
 DB 398 TQHDYLDHSDIIGMTREGTEKPGSLAALITDGPCKSMYVKGQHGAKVYDYLDGNNS 457
 QY 284 DELGKTFIDYFPEYQTEDGYVS-----DKRSITGSKSYETRPMDLTGOFI 328
 DB 458 DTVITINS--DGWGEFVNGSVSVWVPRKTTVTIARPTTRPW--TGFEV 504

RESULT 9
 AA97547
 ID AA97547 standard: Protein; 515 AA.
 XX
 XX AA97547;
 AC
 XX
 DT 12-FEB-2001 (first entry)
 XX
 DE B. stearothermophilus termamyl-like alpha amylase.
 XX
 KW Termamyl-like alpha-amylase; variant; starch liquefaction; fuel;
 KW detergent composition; laundry cleaning composition; ethanol production;

KW dish washing cleaning composition; hard surface cleaning composition;
 KW industrial ethanol production; textile desizing.
 XX
 OS Bacillus stearothermophilus.
 XX
 PN WO200060059-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 28-MAR-2000; 2000WO-DK00148.
 XX
 PR 30-MAR-1999; 99DK-0000437.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Andersen C, Jorgensen CT, Bisgard-Frantzen H, Svendsen A;
 PI Kjaerulff S;
 DR WPI: 2001-015656/02.
 DR N-PSDB: AAA37851.
 XX
 PS Claim 17: Page 69-70; 78pp: English.

XX This sequence represents a termamyl-like alpha amylase.
 CC The invention relates to a variant (I) of parent Termamyl-like
 CC alpha-amylase comprising alteration at one or more of the positions
 CC M13, G48, T49, S50, O51, A52, D53, V54, G57, G108, A111, S168 and
 CC M197. The alterations in (I) are independently an insertion of an amino
 CC acid downstream of the amino acid which occupies the position or deletion
 CC or substitution of the amino acid which occupies the position with a
 CC different amino acid. The variant has alpha-amylase activity. (I) or
 CC compositions containing it are useful in starch liquefaction, in
 CC detergent composition such as laundry, dish washing and hard surface
 CC cleaning compositions, ethanol production such as fuel, drinking and
 CC industrial ethanol production, desizing of textiles, fabrics or garments.
 CC (I) exhibits a reduced capability of cleaving a substrate close to the
 CC branching point, and further exhibits improved substrate specificity
 CC and/or improved specific activity relative to the parent alpha-amylase.
 XX

Query Match 6.4%; Score 110.5; DB 22; Length 515;
 Best Local Similarity 19.4%; Pred. No. 0.028; Mismatches 120; Indels 113; Gaps 16;
 Matches 68; Conservative 50; Mismatches 120; Indels 113; Gaps 16;
 QY 50 SYKEFAESALEETRRRIYLEPHSHLOEKYSMTGLDQKTELQMLAFKSKIRLLRSMAIEN 109
 DB 195 NYDIYMTADLDMDHDEVTTELKNMGKWTNTNIDG-----FRIDAVKH 238
 QY 110 ETMOVF-DPAGMEPEYDPIFCANFETSTNVIYVLDNPLHQLTDQD-----Y 159
 DB 239 IKFSFPDMLSYRSQTGKPLFTVGEYWS-----YDINKLHNTTKDGTMSLFDAPL 291
 QY 160 QDKYY-----NKIMSIYHKAETP-----PWGKRLTG 186
 DB 292 HNKFTYASKSGAFDMRTLMNTLMKDQPTLAVTFVDNHDTEPGALQSMVDPW----- 345
 QY 187 ESIKFSPPLVMTTRSSSEKHKALFSAFLEYQAMLEMTIQVREMEPSHVRANCEA-- 244
 DB 346 -----FKPLA-YAFILTRQEGYPCVF--YGDYIGIPQVNIPIPSLKSIDPLILARRDYAYG 397
 QY 245 -QHKYLT-----WRAQ--KDPGHGLKRL-----VGEAKAKELLRLDFLFCV 283
 DB 398 TQHDYLDHSDIIGMTREGTEKPGSLAALITDGPCKSMYVKGQHGAKVYDYLDGNNS 457
 QY 284 DELGKTFIDYFPEYQTEDGYVS-----DKRSITGSKSYETRPMDLTGOFI 328
 DB 458 DTVITINS--DGWGEFVNGSVSVWVPRKTTVTIARPTTRPW--TGFEV 504

```

RESULT 10
AB076588
ID ABB76588 standard; Protein: 515 AA.
XX
AC ABB76588;
XX
DT 19-AUG-2002 (first entry)
XX
DE Termamyl-like-alpha-amylase #3.
XX
KW Termamyl; alpha amylase; starch liquefaction; ethanol production;
textile desizing; detergent; enzyme.
XX
OS Bacillus stearothermophilus.
XX
PN WO200210355-A2.
XX
PD 07-FEB-2002.
XX
PE 12-JUL-2001; 2001WO-DK00488.
XX
PR 01-AUG-2000; 2000DK-0001160.
PR 12-SEP-2000; 2000DK-0001354.
PR 10-NOV-2000; 2000DK-0001687.
PR 26-APR-2001; 2001DK-0000655.
XX
PA (NOVO ) NOVOZYMES AS.
XX
PI Thilsted T, Kjaerulff S, Andersen C, Fuglsang CC;
XX
DR WPI; 2002-280633/32.
XX
DR N-PSDB; ABL96209.
XX
PT Variant of parent Termamyl-like alpha amylase, useful in detergent
compositions, for starch liquefaction, ethanol production, washing
and/or dish washing, and textile desizing
XX
PS Claim 4; Fig 5; 90pp; English.
XX
CC This invention relates to variants of a parent Termamyl-like
alpha-amylases. These are used for starch liquefaction, ethanol
production, detergent, and textile desizing. The amylases have altered
stability, particularly at high temperatures from 70-120plusoc and
low pH in the range from pH 4.0-6.0. The present sequence is a
termamyl-like-alpha-amylase.
XX
SQ Sequence 515 AA:
Query Match 6.4%; Score 110.5; DB 23; Length 515;
Best Local Similarity 19.4%; Pred. No. 0.028;
Matches 68; Conservative 50; Mismatches 120; Indels 113; Gaps 16;
XX
YY 50 SYKEFESALEETRRKRIIVLEPSHLOEKYSMTGLDQTELOMAFSSKIRLRSAIEN 109
DB 195 NYIYLMATLDDMDHPEVYVELKMKWKYVNTINIDG-----FRIDAVKH 238
YY 110 ETMQVF-DFAGMEPEYDPIPCANFTSTNNIVLNLPHQLTDQTD-----Y 159
DB 239 IKRSFPPDMLSYRSQTKPLFTVGEYWS-----YDINKLNIYTKIDGTMSLDAPL 291
YY 160 QDKYY-----NKIMSIYKRYAETP-----FWGSKLTG 186
DB 292 HNKFFYASKSGAFDMKRTIMTNTLMKDOPLAVTFVDNHDTEPGQALQSWDPW----- 345
YY 187 ESIKFSPPLVMYTRFSSSEKHKALFSAFLFVYQAMLEMTIQVREMEPSHVANCEA-- 244
DB 346 -----FKPLA-YAFILTRQEGIPCVP--TGDYIGIQYNIPLSKSIDPLLARRDYATG 397
YY 245 -OHKYLTLT-----WRAQ-----KDPGHGLRL-----VGEAKAKELLDPLFNGV 283
DB 398 TQHDYDHSDDIIGMTREGTEKRGSGIALALITDPGSGSKRMVYQKQHAGKVFIDLGNRS 457

```

```

YY 284 DELGKTETIDYEPPEYQTEDGTVS-----DKRSITGSKSYENRPMDLGGQFI 328
DB 458 DTVITNS--DGMGEFKVNGSGSVWVPKRKTTSIARIPTTRPW--TGEFV 504
RESULT 11
AB069335
ID ABB06935 standard; Protein: 515 AA.
XX
AC ABB06935;
XX
DT 19-JUN-2002 (first entry)
XX
DE B. stearothermophilus termamyl-like alpha-amylase protein SEQ ID NO:6.
XX
KW Bacillus; termamyl-like alpha-amylase; alpha-amylase; EC 3.2.1.1;
variant; mutant; enzyme; protein co-ordinate data; cleaning; detergent;
washing; sweetener; ethanol; starch.
XX
OS Bacillus stearothermophilus.
XX
PN WO200166712-A2.
XX
PD 13-SEP-2001.
XX
PE 07-MAR-2001; 2001WO-DK00144.
XX
PR 08-MAR-2000; 2000DK-0000376.
PR 15-MAR-2000; 2000US-189857P.
PR 23-FEB-2001; 2001DK-0000303.
PR 26-FEB-2001; 2001US-271382P.
XX
PA (NOVO ) NOVOZYMES AS.
XX
PI Andersen C, Borchert TV, Nielsen BR;
XX
DR WPI; 2002-239612/29.
XX
DR N-PSDB; ABL50566.
XX
PT Novel variant of parent termamyl-like alpha-amylase useful as a
component in washing and dishwashing compositions, for textile
desizing, for starch liquefaction, and for producing sweeteners and
ethanol from starch
XX
PS Claim 8; Page 138-139; 153pp; English.
XX
CC The present invention describes a variant of a parent termamyl-like
alpha-amylase (EC 3.2.1.1) (I) comprising an alteration at one or more
positions of a group of 31 possible amino acid positions. The alteration
in (I) may be at Arg28, Arg118, Asn174, Arg181, Gly182, Gly184,
Gly186, Trp189, Asn195, Met202, Tyr298, Asn299, Lys302, Ser303,
Asn306, Arg310, Asn314, Arg320, His324, Glu345, Tyr396, Arg400,
Asn445, Lys446, Arg449, Arg458, Asn471, or Asn484. (I) can be used for
washing and/or dishwashing, textile desizing, and starch liquefaction.
CC (I) is useful as a component in hard surface cleaning detergent
composition, and for producing sweeteners and ethanol from starch.
CC (I) has altered solubility, preferably increased solubility, in
particular under washing, dish washing or hard surface cleaning
conditions. The present sequence represents a Bacillus stearothermophilus
termamyl-like alpha-amylase which is used in the exemplification of the
present invention.
XX
SQ Sequence 515 AA:
Query Match 6.4%; Score 110.5; DB 23; Length 515;
Best Local Similarity 19.4%; Pred. No. 0.028;
Matches 68; Conservative 50; Mismatches 120; Indels 113; Gaps 16;
XX
YY 50 SYKEFESALEETRRKRIIVLEPSHLOEKYSMTGLDQTELOMAFSSKIRLRSAIEN 109
DB 195 NYIYLMATLDDMDHPEVYVELKMKWKYVNTINIDG-----FRIDAVKH 238
YY 110 ETMQVF-DFAGMEPEYDPIPCANFTSTNNIVLNLPHQLTDQTD-----Y 159

```

```

Db      239 IKFSEFPDMLSYRSOTGKPLETVGEYWS-----YDINKLHNITKTGDMSLFDAPL 291
QY      160 QDKYY-----NKINSIYKHAETP-----PMGSKLNG 186
Db      232 HNKFTASKSGAFDMRTLMNTLMKDOPTLAVTFVDNHDTEPGALQSWVDPW----- 345
QY      187 ESKEFSPPLVMTFRSSSEKHKALFSAFLEYQAMLEMTIOVREMEPSHYRANCEA-- 244
Db      346 -----FKPLA-YAFILTROEGPCVF--YGDYIGIPQYNIPSLKSKIDPLIARDYAYG 397
QY      245 -QHKYLT-----WRAQ---KDPGHGLKRL-----VGEAKAKELLRDLFLNGV 283
Db      398 TQHDYLDHSDITIGMTREGGTEKPGSGLAALITDGGGSKMMYVGKOHAKVYDLDGNRS 457
QY      284 DELGKRTFDYFPEYQTEGTVS-----DKRSIIGKSYETRPMDLTGQFI 328
Db      458 DTVTINS--DGMGEFKVNGSVSWVPRKTYVTIARPTTRPW--TGSEFV 504

```

RESULT 12

AA047852
ID AAB47852 standard; Protein; 515 AA.

AA047852;
AC AAB47852;

DT 02-APR-2002 (first entry)

DE Bacillus alpha amylase BSG.

XX Alpha amylase; Bacillus; Termamyl-Like; maltodextrin; glucose syrup;
KW starch; food; feed; pharmaceutical; confectionery; candy;
KM isotonic drink; bakery; cereal bar; ice cream; coffee whitener;
KW salad dressing; cured meat; fermented meat; spice.

OS Bacillus stearotheophilus.

PN WO200196537-A2.

PD 20-DEC-2001.

PF 13-JUN-2001; 2001WO-DK00404.

PR 14-JUN-2000; 2000DK-0000917.

PR 20-JUN-2000; 2000US-212852P.

XX (NOVO) NOVOZYMES AS.

PI Nielsen BR, Welbye M;

DR WPI: 2002-098064/13.

DR N-PSDB: AA172213.

PT New modified alpha-amylase derived from the genus Bacillus and/or is a
PT Termamyl-Like alpha-amylase, which has been pre-oxidized for producing
PT maltodextrin or glucose syrup

PS Claim 5; Page 32-33; 47pp; English.

CC The sequences given in AAB47850-56 show modified alpha-amylases
CC derived from the genus Bacillus. These alpha amylases are Termamyl-
CC like alpha-amylase and they have been pre-oxidized. The alpha amylase
CC is useful for producing a maltodextrin or glucose syrup, by treating
CC starch with a pre-oxidized alpha-amylase until a product with a
CC DE between 5-45 has been provided and/or until a product with a
CC molecular weight of between 5-30 kda has been provided. The product
CC comprises a maltodextrin with a DE of 18.5 and/or a maltodextrin with a
CC molecular weight of 14-16 kda. The alpha amylase is useful for producing
CC a maltodextrin or glucose syrup, where the glucose syrup is useful as an
CC ingredient in food, feed or pharmaceuticals. Glucose syrup is useful
CC in confectionery such as candies, beverages such as isotonic drinks,
CC bakery such as cereal bars, dairy and ice cream such as coffee
CC whiteners, conventional foods such as salad dressings, and food

CC ingredients and preparations such as cured meat, fermented meat, spices
CC and seasoning encapsulated flavours.

SQ Sequence 515 AA;

Query Match 6.4%; Score 110.5; DB 23; Length 515;
Best Local Similarity 19.4%; Pred. No. 0.028;
Matches 68; Conservative 50; Mismatches 120; Indels 113; Gaps 16;

```

QY      50 SKEFSALEETRRRIYLEPSHLOEKSSMTGLDGTLOMLAFKSSKIRLLSMALEN 109
Db      195 NYDYLMYADLDHDHDEVVTELKMKGVNTNTIDG-----FRIDAVKH 238
QY      110 ERMQVE-PRAGMEDEYDIPICANFETNNNIWLDNPLHQLDQD-----Y 159
Db      239 IKFSEFPDMLSYRSOTGKPLETVGEYWS-----YDINKLHNITKTGDMSLFDAPL 291
QY      160 QDKYY-----NKINSIYKHAETP-----PMGSKLNG 186
Db      292 HNKFTASKSGAFDMRTLMNTLMKDOPTLAVTFVDNHDTEPGALQSWVDPW----- 345
QY      187 ESKEFSPPLVMTFRSSSEKHKALFSAFLEYQAMLEMTIOVREMEPSHYRANCEA-- 244
Db      346 -----FKPLA-YAFILTROEGPCVF--YGDYIGIPQYNIPSLKSKIDPLIARDYAYG 397
QY      245 -QHKYLT-----WRAQ---KDPGHGLKRL-----VGEAKAKELLRDLFLNGV 283
Db      398 TQHDYLDHSDITIGMTREGGTEKPGSGLAALITDGGGSKMMYVGKOHAKVYDLDGNRS 457
QY      284 DELGKRTFDYFPEYQTEGTVS-----DKRSIIGKSYETRPMDLTGQFI 328
Db      458 DTVTINS--DGMGEFKVNGSVSWVPRKTYVTIARPTTRPW--TGSEFV 504

```

RESULT 13

AA012151
ID AA012151 standard; Protein; 515 AA.

AA012151;
AC AA012151;

DT 09-APR-2002 (first entry)

DE Bacillus TERMAMYL-Like alpha-amylase BSG.

XX TERMAMYL; alpha-amylase; detergent; dishwashing; textile desizing;
KW starch liquefaction; ethanol production; hard surface cleaner;
KM sweetener; amylopectin; limit dextrin; NOVAMYL; BSG.

OS Bacillus stearotheophilus.

PN WO200188107-A2.

PD 22-NOV-2001.

PF 10-MAY-2001; 2001WO-DK00323.

PR 12-MAY-2000; 2000DK-0000779.

XX (NOVO) NOVOZYMES AS.

PI Svendsen A, Jorgensen CT, Nielsen BR;

DR WPI: 2002-106123/14.

DR N-PSDB: AAS20024.

PT New variant of parent Termamyl-Like alpha-amylase for use as a
PT component in washing and dishwashing compositions, for textile
PT desizing, for starch liquefaction, and for producing sweeteners and
PT ethanol from starch

PS Claim 5; Fig 1; 84pp; English.

CC The invention relates to a variant of parent TERMAMYL-Like alpha-

